REGULATING PRIVACY ONLINE: AN ECONOMIC EVALUATION OF THE GDPR*

Samuel G. Goldberg,
Garrett A. Johnson,
Scott K. Shriver

November 16, 2022

Abstract

Modern websites rely on personal data to measure and improve their performance and to market to consumers. The European Union's General Data Protection Regulation (GDPR) limited access to such personal data, with the goal of protecting consumer privacy. We examine the GDPR's impact on website pageviews and revenue for 1,084 diverse online firms using data from Adobe's website analytics platform. Among EU users, we find a reduction of approximately 12% in both website pageviews and e-commerce revenue, as *recorded* by the platform after the GDPR's enforcement deadline. We find evidence that the GDPR both reduced data recording and harmed real economic outcomes, and we derive bounds for the relative contribution of each explanation.

JEL Codes: L51, L86, M31, M38

[*] Corresponding Author: Samuel Goldberg, Stanford Institute for Economic Policy Research, Stanford University, 366 Galvez Street, Stanford CA 94305. <Sggold@stanford.edu> ; Johnson: Questrom School of Business, Boston University <garjoh@bu.edu>; Shriver: Leeds School of Business, University of Colorado – Boulder <Scott.Shriver@colorado.edu>. The authors would like to thank Adobe for providing the data, Taylor Schreiner for his support within Adobe, Michail Batikas, Avi Goldfarb, Anita Rao, and Catherine Tucker for helpful discussions, Dirk Bergemann, Andrey Fradkin, Brett Gordon, Laura Kornish, Tesary Lin and Brad Shapiro for comments, as well as seminar participants at Boston University, Digital Economics Paris 2022, FTC Microeconomics 2021, George Mason University School of Business, Harvard Business School, INSEAD, Johnson Graduate School of Management, Joint Digital Economics Seminars, Kellogg School of Management, Lehigh University, Marketing Science 2019, NBER Digitization Summer Institute 2019, and the ZEW Conference on the Economics of ICT 2021. The research was conducted while Samuel Goldberg was an employee of Adobe and partially funded by MSI research grant #4000783. All errors are our own.

1 Introduction

Personal data is a fundamental input of the modern digital economy. Firms use personal data for a variety of activities including product recommendations, advertising, and pricing. Concerns over invasive and opaque data collection practices have fueled a political debate about the consumer's right to privacy and how best to regulate it. The European Union's (EU) General Data Protection Regulation (GDPR) seeks to protect individual privacy by reigning in firms' use of personal data. The GDPR is one of the most significant privacy regulations in effect today and has served as a model for subsequent privacy regulations such as California's Consumer Privacy Act. While privacy may be valuable to individuals, the GDPR's limitations on the use of personal data may harm online firms that represent 4% of the EU's economy and 10% of its retail sales.[1] Protecting privacy may therefore be costly. This paper presents a broad-based empirical study of the GDPR's economic impact on online firms. We examine the effect of the GDPR on site traffic—a measure of site health and its capacity to generate advertising revenue—as well as site revenue arising from e-commerce sales.

A key contribution of this paper is to distinguish between the GDPR's impact on real economic activity and the process of recording economic outcomes by online firms. Extant policy research examines how regulation affects economic outcomes, but seldom contends with policies that also affect how those outcomes are recorded. Privacy policy creates missingness by construction because individuals can forbid their data from being recorded or shared. Under the GDPR, *recorded economic outcomes* may fall because some individuals do not consent to data sharing. Indeed, changes to when and what data are recorded is a primary goal of the GDPR. At the same time, a decline in recorded outcomes could reflect a decline in *real economic outcomes*, for instance, because the regulation restricts personalized marketing. Privacy regulation thus creates an inference problem: data protection can both impact economic outcomes *and* obscure the observation of economic outcomes. Consequently, policymakers need to distinguish between the real and recording effects of privacy regulation in order to evaluate it. Further, both real and recording effects are policy relevant. Real effects are helpful for understanding the impact of privacy regulation on firms, while changes in data recording may signal privacy-related welfare gains for consumers. Though regulators can observe aggregate firm economic activity, they may require more frequent and detailed data to detect and diagnose policy effects—particularly when such effects are small. We utilize such data to show that the GDPR reduces recorded outcomes by 12% and then we bound the contributions of the real and recording effects of the GDPR.

[1] https://www.retailresearch.org/online-retail.html

We use proprietary data from Adobe Analytics to understand the implications of the GDPR for the online economy. Our data consist of recorded outcomes from over 1,000 websites from such diverse industries as media, travel, beauty, health, and retail. We observe 77 of the top 1,000 websites globally as well as over 700 sites below the top 100,000. The data separate the web traffic by EU residents who are protected by the GDPR. Our analysis focuses on two key metrics of the online economy: pageviews and revenue arising from e-commerce sales. We observe over 4.4 billion weekly pageviews and about \$0.75 billion in weekly e-commerce revenue from EU users alone—representing 12% of total European e-commerce. For advertising-supported websites, pageviews indicate a site's supply of ads and ad revenue. E-commerce revenue contributes to overall economic activity as 17% of EU enterprises sold online in 2018 (EuroStat, 2020). The data also identify the channels (e.g. search engine, email, display ad) through which users arrive at the website, which we use to examine how the GDPR affected online marketing to EU users.

Like past economic studies of privacy regulation (e.g., Goldfarb and Tucker, 2011; Miller and Tucker, 2009), we use the timing of regulatory enforcement as an event study. The GDPR affects firms at the same time, so selecting an appropriate control group is challenging. Our preferred control group is the same set of sites in the previous year. This control group then captures the seasonal pattern of EU user traffic, which is specific to these firms. Our primary analyses use a panel differences approach (similar in spirit to a differences-in-differences model) to identify the impact of the GDPR on recorded pageviews and recorded revenue.[2] We estimate that recorded pageviews fall 11.7% post-GDPR, on average across all sites, or 15,043 pageviews per week for the median site. Among e-commerce sites, recorded revenue falls 13.3%, or \$9,227 per week for the median site. While the above results suggest that the GDPR has changed recorded online outcomes, they do not disentangle the role of data recording from real economic harms.

The GDPR can impact recorded web outcomes through two main mechanisms. First, European privacy regulators state that sites must obtain user consent for collecting site analytics data (Article 29 Data Protection Working Party, 2012; Information Commissioners Office, 2019; Data Protection Commission, 2020a), such that site visitation data from non-consenting users is unrecorded. The most adopted consent management platform indicates that over 90% of users provide consent (Quantcast, 2018). If up to 10% of users do not provide consent, this could explain some of the 11.7% reduction in recorded pageviews. Second, privacy regulation may affect sites' real outcomes, for instance by limiting their

[2] Formally, differences-in-differences requires contemporaneous observation of cross-sectional units that are not exposed to the treatment of interest. The GDPR's scope and coordinated rollout effectively precludes the construction of a true control group, since websites serving EU citizens (virtually all major sites) are subject to the GDPR.

marketing ability. Privacy regulation has been documented to negatively impact marketing effectiveness, both in the context of the EU's 2003 ePrivacy Directive (Goldfarb and Tucker, 2011) and the GDPR (Godinho de Matos and Adjerid, 2021; Aridor et al., 2020). Marketing may be less effective particularly on channels like email and display advertising that rely on personal data. This marketing mechanism would lead to a decrease in real site outcomes. Finally, though some firms may stop sharing data with Adobe to conform with the GDPR's principle of data minimization, we rule this mechanism out by excluding such firms from our analysis.

We quantify the relative contributions of the consent and real effect mechanisms (for example, such as marketing effectiveness). Our framework exploits the fact that consenting users are favorably selected to construct bounds on the role of consent after the GDPR. In particular, we allow the post-GDPR reduction on the extensive margin (recorded visits) to be a product of both the consent and real effects. However, we constrain the post-GDPR improvement on the intensive margin (pageviews per recorded site visit) to arise from the consent effect alone. We then recover the consent effect lower bound (and the implied upper bound on the real effect) by exploiting the natural lower bound of one pageview per visit for unobserved, non-consenting users.

We estimate non-consent rates are between 4.0% and 12.8% of site visits in our full sample and between 3.6% and 13.7% for our e-commerce site sample, which aligns with the consent rates in Quantcast (2018). In particular, consent explains a minimum of 4.7% of our GDPR estimate for recorded pageviews. Given the policy's explicit objective of limiting data collection, the non-consent rate is an informative metric for evaluating the GDPR. Our estimates suggest that a non-negligible portion of consumers are benefitting from the ability to opt-out of data collection.

The remainder of the GDPR estimate could be due a real effect on EU web traffic, though conservative estimates suggest a small and negative marketing-related effect. In particular, we use data on the user's arrival channel to look for evidence that marketing has been impacted by the GDPR. We find that the GDPR had a greater impact on users arriving through data-intensive marketing channels like display and email advertising. Net of the consent effect, we calculate that marketing mechanisms reduce real pageviews at least 0.5% per week. For e-commerce revenue, marketing mechanisms reduce real revenue at least 0.4% per week. Thus, the real effect of the GDPR through the display and email channels alone represent 7.0% of the GDPR estimate on pageviews and 4.6% of the e-commerce revenue estimates.

We show the GDPR has differential effects on sites by size and regulatory strictness. Extant research highlights anti-competitive effects of the GDPR (Johnson et al., 2022; Peukert et al., 2022). We find smaller e-commerce sites see almost twice the decline in recorded revenue (-16.7%) than larger sites (-7.9%). We show this decline arises from smaller e-commerce sites having a harder time obtaining consent, which validates a key theoretical prediction in Campbell et al. (2015). We also consider the moderating effect of regulatory strictness. The GDPR harmonized regulation within the EU, but enforcement is at the country level where regulator resources vary. We show that site beliefs about data protection regulatory strictness—as proxied by a EU survey of firms—moderate the effect of the GDPR. Specifically, a one standard deviation increase in the regulatory strictness index corresponds to 2.0% lower recorded pageviews and 2.1% lower recorded revenue. This finding aligns with other research showing a correlation between regulatory strictness and the economic effects of the GDPR (Jia et al., 2021; Johnson et al., 2022).

Our work expands the literature on the economic implications of privacy regulation. Pre-GDPR studies of privacy regulation, for instance, show that privacy laws can slow technology diffusion (Miller and Tucker, 2009; Adjerid et al., 2016) and that the EU's 2009 e-Privacy Directive reduced advertising effectiveness (Goldfarb and Tucker, 2011). Prasad and Perez (2020) and Johnson (2022) summarize the economics literature on the GDPR. Jia et al. (2020, 2021) note that venture capital investment falls in the EU post-GDPR. Koski and Valmari (2020) and Chen (2022) find evidence that the GDPR reduced firm profits using accounting data. Janssen et al. (2022) document reductions in Android mobile app availability and entry. Theoretical research suggests that the GDPR's privacy rights can strengthen consumers at the expense of firms (Ke and Sudhir, 2020). Computer science research shows that websites use fewer third-party vendors—which often rely on personal information—after the GDPR (e.g. Libert et al. 2018; Sørensen and Kosta 2019). Johnson et al. (2022) and Peukert et al. (2022) show that this reduction in third-party vendors favored large vendors and thereby increased market concentration. However, Zhuo et al. (2021) show no effect of the GDPR at the Internet infrastructure level.

Aridor et al. (2020), Lefrere et al. (2020), Zhao et al. (2021), and Schmitt et al. (2021) are closest to our work in that they also examine how the GDPR affects websites. Aridor et al. (2020) uses analogous data as they examine traffic on travel sites as recorded by a third-party intermediary that sells travel advertising. Their findings complement ours in that both papers estimate similar-sized effects on recorded site traffic and find that recorded outcomes are favorably selected post-GDPR. Aridor et al. (2020) attribute the GDPR effect solely to consent, whereas we explore alternative mechanisms by leveraging our greater cross-section of sites as well as data on visits by marketing channel. Aridor et al. (2020) instead use their ad revenue and user scoring data to show that, though the GDPR hinders firm's marketing, this is somewhat offset by the favorable selection of consenting users.

Zhao et al. (2021) instead examine browsing behavior using an opt-in panel of web users and highlight an increase in search intensity due to the GDPR. Using an individual-level panel eliminates the missing data problem created by non-consenting users in firm-aggregate panel data, but selection concerns in turn create challenges for population-level inference. Lefrere et al. (2020) and Schmitt et al. (2021) instead rely on data from SimilarWeb and Alexa respectively, which apparently combine user panel data with site analytics data to estimate site traffic. Lefrere et al. (2020) find no change in pageviews and instead find a small decrease in pageviews per user on EU sites due to the GDPR. Schmitt et al. (2021) instead find a 5% reduction in estimated site traffic.

This paper is also related to a growing literature on the use of analytics and internet data to measure the economy. Early studies focused on the value of analytics and data-driven decision-making to firms (Bresnahan et al., 2002; Brynjolfsson et al., 2011). Recent work by Berman and Israeli (2022) shows positive returns to firms that adopt a data analytics platform. A handful of studies use web analytics data—including price and transaction data—to measure the online economy. Goolsbee and Klenow (2018) measure price inflation online using data from e-commerce sites from our same provider, Adobe Analytics. Other work estimates the value of the internet and e-commerce (Cavallo and Rigobon, 2016; Dolfen et al., 2019; Brynjolfsson and Oh, 2012). Budak et al. (2016) is closest to our study as they use browsing history to assess how e-commerce sites depend on online advertising to generate traffic. We contribute to this literature by tackling missingness in web analytics data due to user consent, in the spirit of Manski (2005).

The paper proceeds as follows. Section 2 provides an overview of the GDPR. In Section 3, we explain how the GDPR impacts recorded outcomes and describe our data. In Section 4, we present our empirical approach, our main GDPR effect estimates, and explore heterogeneity between firms. Section 5 devises bounds for the contributions of the real and consent effects of the GDPR. Section 6 concludes.

2 GDPR background

The European Union (EU) passed the General Data Protection Regulation (GDPR) in April of 2016 and enforcement began on May 25, 2018, giving firms two years to prepare. The GDPR protects the collection, processing, and use of personal information of EU residents. The GDPR expands the definition of personal data to include individual-specific data like cookie identifiers and IP addresses (Article 4(1)). The GDPR covers both EU firms as well as foreign firms that target EU residents. Fines in the case of non-compliance can reach the larger of 20 million euros or 4% of global revenue. The GDPR regulators split enforcement between country-specific regulators and an EU-level regulator.

Data minimization is a guiding principle of the GDPR (Article 5(1c)): firms must limit the personal data they collect and use. The GDPR accords data rights to EU residents including the right to access, correct, erase, and port their personal data. Firms that process personal data must invest in systems and processes to fulfill these rights-based responsibilities. These firms are also required to audit internal data processes and to appoint a data protection officer to oversee compliance activities. Firms must encrypt and anonymize personal data as well as promptly notify both the regulator and affected individuals after a data breach. These obligations impose significant compliance costs on firms. Many firms spent over 10 million dollars annually to comply with the law and many came into compliance after May 25, 2018 (PricewaterhouseCoopers, 2018). The GDPR requirements incentivize firms to minimize their data processing, including for activities like marketing.

The GPDR's Article 6(1) lays out legal bases under which a firm may process personal data. EU regulators clarified that individual consent is the most appropriate basis for websites and their vendors to process user data—including for web analytics purposes (Article 29 Data Protection Working Party, 2012; Information Commissioners Office, 2019; Data Protection Commission, 2020a).[3] Under the GDPR, valid consent must be affirmative (no pre-checked boxes), freely given, granular to the purpose of processing (e.g. website analytics, behavioral advertising), and must list all third parties who process the data (European Data Protection Board, 2020).

The reality of GDPR as implemented by firms can differ from the regulation as written. The GPDR is a multi-faceted regulation, and firms were initially uncertain as to how to comply. Foreign firms, in particular, were unfamiliar with the regulation's legal context, which compounded their confusion (Jones and Kaminski, 2020). In surveys, a minority of firms described themselves as GDPR-compliant as of June 2018, though EU firms were ahead of their American counterparts (TrustArc, 2018). Adding to this uncertainty, EU regulators continued to issue new guidance after the enforcement deadline and moved slowly to enforce the law. The European Commission (2019) later acknowledged that the GDPR was under-enforced in its first year. Investigations during that time reveal that most websites did not wait for consent before interacting with third party domains or setting identifier cookies (Sanchez-Rola et al., 2019; Johnson et al., 2022). Even in 2019, the Irish privacy regulator

[3] European regulators have long signaled some openness to web analytics that exclusively use the site's data (i.e. using first party cookies), but concern about web analytics vendors that combine user data across sites (i.e. using third party cookies) (Article 29 Data Protection Working Party, 2012). In 2020, the French regulator opened the door for sites to collect web analytics data without opt-in consent under the French Data Protection Act, provided that this data is not combined with off-site data (Commission Nationale de l'Informatique et des Libertés, 2020b).

found that almost all sites it investigated were still setting cookies upon the user's arrival (Data Protection Commission, 2020b). EU regulators objected to these practices (Commission Nationale de l'Informatique et des Libertés, 2020a; Data Protection Commission, 2020b), but did not fine websites until December 2020.

We focus on the GDPR's effects on two main channels: consent and personalized marketing. If a user refuses to consent, websites can not send user data to Adobe Analytics. Consent rates therefore have important consequences for the data recorded by Adobe. In addition, many websites employ personalized marketing to increase traffic. By limiting the use of personal data in marketing, the GDPR may reduce the real traffic to websites. We also allow for possibility that the GDPR reduced real visits to websites more generally: for instance, ubiquitous consent popups may create a distaste for visiting websites. However, our analysis pays special attention to the personalized marketing explanation because our data can speak to this.

Consent Rather than follow the GDPR's *de jure* opt-in consent requirements, most websites instead followed a *de facto* opt-out approach when obtaining user consent. Contemporary investigations by researchers and EU privacy regulators confirmed that most sites improperly relied on opt-out consent (Autoriteit Persoonsgegevens, 2019; Utz et al., 2019). For instance, many sites employed consent dialogs that invited the user to "agree" to continue or click "more options" to opt out of specific data processing on a second page. This approach to consent ensures high consent rates, and the largest GDPR consent management platform reported average consent rates in excess of 90% (Quantcast, 2018). By contrast, sites that followed a strict opt-in approach have much lower consent rates. In this case, the British privacy regulator's site and the Dutch public broadcaster's sites report consent rates of 10% or lower (Information Commissioner's Office, 2019; Snelders et al., 2020). Note that the British privacy regulator also provided site analytics data showing a commensurate drop in recorded traffic of about 90%. Based on the Quantcast (2018) figure, we therefore expect to see a reduction up to about 10% in recorded traffic data due to non-consenting users. In Section 5.2, we estimate bounds on the average non-consent rate in our data to contribute to the policy discussion on *de facto* consent under the GDPR.

Personalized marketing The GDPR raised the legal risk and logistical cost associated with personal data processing. Firms may respond by reducing their investment in marketing channels like e-mail and online display advertising. Those media rely on personal data in the form of e-mail addresses and cookie-based identifiers, respectively. The GDPR was expected to reduce email marketing because firms would reduce or even delete their customer email lists. For instance, many firms sent permissioning emails (Godinho de Matos and Adjerid, 2021) prior to the GDPR seeking opt-in consent to continue emailing users, then dropped non-consenting users from email marketing lists. In online display advertising, Johnson et al. (2022) show moderate and persistent reductions in related ad technology vendor use by websites after the GDPR in 2018. In addition, firms report high costs of complying with the GDPR, which may divert funds from marketing if firms view this expense as discretionary. Thus, the quantity and effectiveness of personalized marketing channels may fall post-GDPR. Nevertheless, Budak et al. (2016) conservatively estimate that only around 10% of site traffic originates from email or display ad clicks. In Sections 4.3.1 and 5.3, we provide empirical evidence that the GDPR may impact traffic to websites through personalized marketing channels.

3 Data

Our study uses proprietary data from the Adobe Analytics platform to study the GDPR. Section 3.1 overviews the services and data that web analytics vendors provide. Section 3.2 explains how GDPR affects the data that web analytics platforms record. Section 3.3 describes our panel of online firms and how we construct the panel. Finally, Section 3.4 provides summary statistics for our key outcomes.

3.1 Web analytics overview

Online firms use web analytics tools to understand the characteristics and behaviors of their site visitors. Our data is provided by Adobe Analytics, a leading web analytics vendor (Forrester, 2017). Web analytics vendors like Adobe Analytics provide technology for websites to track users who browse their sites. In order to implement Adobe Analytics, the firm adds code to their website which sends a data-rich ping to Adobe's servers whenever a user visits. These pings contain a unique user ID, website ID, webpage information, and generate a timestamp.[4] Adobe then aggregates the ping data into an *analytics dashboard*. Analytics dashboards are the primary unit of analysis in this paper. Dashboards reveal traffic and revenue performance aggregated over time and broken down, for instance, to the user's country of residence.[5] Online firms find these dashboards to be a source of customer insight that can increase revenue (Berman and Israeli, 2022). For instance, changes in web analytics performance may alert the firm to opportunities to improve its website design or marketing

[4] We do not observe the unique user ID in our data, but simply aggregated site-level statistics. Thus, we can not track individual user behavior over time.

[5] Refer to the Adobe Analytics (Adobe, 2019) documentation for technical information.

activities. Even the French privacy regulator views site analytics data as “in many cases essential for the proper functioning of the site” (Commission Nationale de l’Informatique et des Libertés, 2020b).

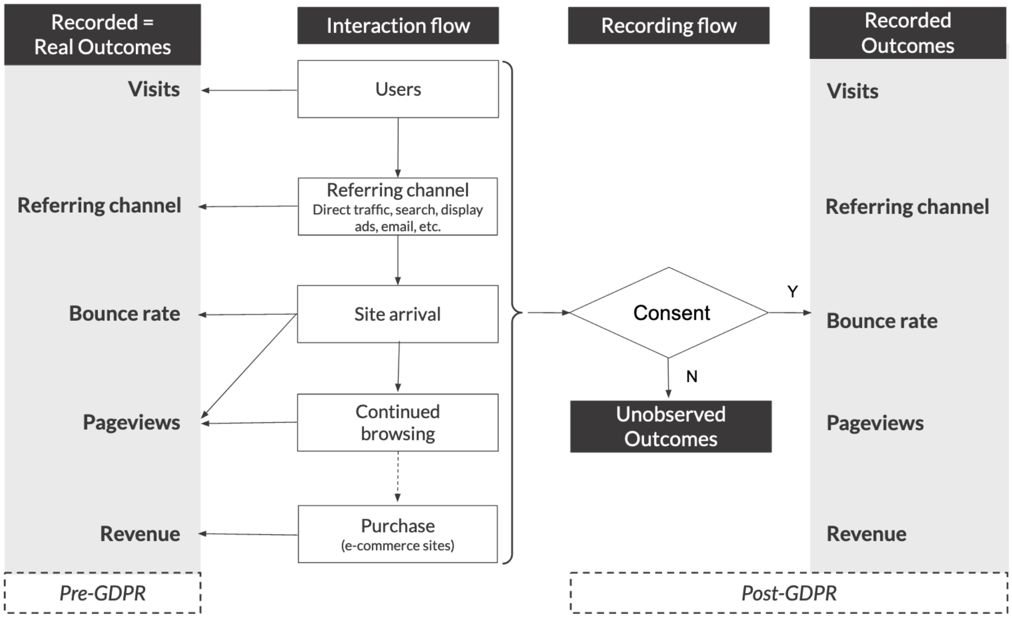
The left-hand side of Figure 1 illustrates how web analytics vendors construct economic variables from website traffic pings. When a user arrives on site, a *site visit* is initiated. As we discuss below, sites receive a signal of how the user arrived at the website: the *referring channel*. Once a user arrives on site, the user may either leave the site or continue to browse more site pages. The former case is a referred to as a *bounce* and the share of such visits is referred to as the *bounce rate*. Whether the user bounces or continues to browse onsite, that user generates site *pageviews*. On e-commerce sites, some users complete a purchase and the web analytics vendor will record the site’s *revenue*.

One advantage of web analytics data is that they describe how users arrive on site, broken out by referring channel. Sites can coordinate with their analytics vendor to infer these channels from referring links. For instance, the site can send a marketing email to a user containing purpose-built hyperlinks. These hyperlink URLs contain not only the landing page address but also the channel (email) through which the user arrives onsite. We observe four referring channels: search, display advertising, email marketing, and social media. Search includes both paid and natural search, and social media includes clicks on both paid and organic social content. Users can also arrive onsite without clicking a link—e.g. by typing the URL into browser navigation bar or by using a bookmark—which is referred to as the direct navigation channel. Referring channel data can also be used to attribute site outcomes (e.g. pageviews, revenue) to these different marketing channels. These data are therefore also referred to as *last-touch attribution* data, because they record the user’s last channel interaction prior to arriving on site. Last-touch attribution is understood to ignore prior channel interactions including those preceding a user’s direct navigation to the site. Nonetheless, changes in a marketing channel’s last-touch attribution share can indicate a change in channel spending and/or channel effectiveness.

3.2 Data recording pre- and post-GDPR

The GDPR affects the share of web analytics data that vendors, like Adobe, record. Figure 1 emphasizes the difference in data recording before the GDPR (left-hand side) and after the GDPR (right-hand side). Before the GDPR, Adobe records all the user traffic onsite. After the GDPR, Adobe only records outcomes from consenting users. If the user consents, Adobe will observe all that user’s outcomes, just as in the pre-GDPR period. If the user revokes consent, Adobe does not record any analytics data associated with that user. As a result,

Figure 1: Web analytics data recording before and after the GDPR



we do not observe any data associated with non-consenting users, including any reference to their visit or site outcomes. We observe *recorded* outcome measures in our data: the consent rate drives a wedge between recorded and real outcomes post-GDPR. To fix ideas, we present a simplified decomposition of our recorded outcomes.[6] Specifically, our recorded outcomes reflect these different components pre- and post-GDPR:

$$\text{Pre: } \textit{Recorded outcomes} = \textit{Real outcomes} \tag{1}$$

$$\text{Post: } \textit{Recorded outcomes} = \textit{Real outcomes} \times \textit{Consent rate} \tag{2}$$

Post-GDPR, recorded outcomes are the product of real outcomes and the consent rate.

The pre- and post-GDPR comparison of recorded outcomes creates an identification challenge. Changes in real outcomes or consent rates below 100% both affect recorded outcomes. However, a measured effect on recorded outcomes does not alone reveal each factor's contribution to the change. We expect that the GDPR impacts both a website's real outcomes *and* reduces data recording through consent. Therefore, we must take care in interpreting changes in recorded outcomes. Our analysis thus proceeds in two stages. In Section 4.2, we measure the effect of the GDPR on recorded outcomes. In Section 5, we propose theory-driven bounds to empirically quantify the contribution of the consent-related and real outcome-related effects of the GDPR.

[6] We consider the more general case where we allow the real outcomes to be correlated with the consent rate in Section 5.

3.3 Sample construction

Our panel data consists of online firms that serve EU residents and that use Adobe Analytics. Our weekly panel data covers 1,084 analytics dashboards for 32 weeks in each of 2017 and 2018.[7] Our unit of analysis is a *dashboard* rather than a website or firm. Firms construct analytics dashboards to provide insight into their online presence, operations, and customers. As such, the dashboard dataset reflects the firm's view of its online presence and can disaggregate large, multi-brand firms. A dashboard may have many-to-many relationships with websites. For example, consider a retailer with both France- and UK-facing sites. The retailer could choose to: 1) combine both sites into one dashboard, 2) create separate dashboards for each site, or 3) create multiple dashboards with duplicate information. In general, firms may elect to organize their websites and/or dashboards by different brands or by customer location. Our primary outcomes of interest are: 1) a pageview: a request for a full webpage document by a site visitor; and 2) revenue: total onsite user spending in US dollars.[8] Both outcomes are economically relevant for online firms. For ad-supported websites, pageviews are proportional to advertising revenue. For e-commerce sites, we directly observe e-commerce revenue.

Our dashboard sample expands on past work using Adobe Analytics data and undergoes several quality control steps. Drawing from the full sample of Adobe's data, we begin with an initial set of about 4,900 dashboards. The initial set requires at least 500 average weekly visits from the EU prior to the GDPR's enforcement and removes dashboards containing duplicate data. The 500-visit threshold ensures EU-relevant data and avoids noisy outcome data due to low EU traffic, while still including long tail sites below the top 100,000 Alexa sites. Critically, we restrict our sample to the 1,425 dashboards that regularly report data: i.e., less than 20% of weeks are missing. This removes many low-quality dashboards including those built for special events or promotions. This step also helps remove firms that join or churn from Adobe, and firms that reorganize their dashboard data structure. Note that some firms may respond to the GDPR's data minimization requirement by dropping Adobe Analytics altogether. However, this filtering step also removes the few dashboards (<0.5%) that turn off reporting post-GPDR. We then remove 328 dashboards that are labeled for testing purposes as well as 164 dashboards whose meta-data omit a web domain name. We remove 118 dashboards whose constituent data elements may have changed as evidenced by unusual growth prior to the GDPR.[9] Finally, we remove dashboards with outlier user

[7] For 2018, our data span the 4th Friday of the year through the week beginning with the 35th Friday of the year (January 28th 2018 to September 7th 2018).

[8] In May 1st, 2018 dollars. Exchange rates are held fixed at May 1st, 2018 rates.

[9] For instance, firms may change the list of domains associated with a dashboard. We remove dashboards that grow more than 170% or fall more than 70% between 2017 and 2018 prior to treatment.

behavior in the pre-GDPR period.[10] Our resulting 1,084 dashboard panel is 99.98% complete for the pageviews outcome and 99.4% complete for the revenue outcome. Our sample's 353 e-commerce dashboards is also much larger than the sample used in Goolsbee and Klenow (2018) because we include non-US firms, examine more recent data, and do not require product-level data.

Our data offer both economic scale and diversity to investigate the GDPR's impact on the EU's digital economy. The final 1,084 dashboard panel includes diverse firms in such industries as finance, health, travel, news/media, and retail. Our sample contains a mix of e-commerce sites, publisher sites, and corporate sites. By observing outcomes by the user's country of origin, we can focus on the EU residents protected by the GDPR. In addition, we gain North American traffic as an alternative control group. Our data contain approximately $0.75 billion per week in European e-commerce revenue, which represents about 12% of total European e-commerce. The data also contain $2.8 billion per week in North American e-commerce revenue—almost a third of North America's estimated spending.[11] The data contain 4.3 billion weekly pageviews from EU users alone. For comparison, Wikipedia saw an estimated 2.6 billion pageviews from EU users in a typical week of 2018.[12]

3.4 Summary statistics

Table 1 illustrates the heterogeneity in our 1,084 dashboard sample. We report pre-GDPR 2018 summary statistics for pageviews, revenue, usage metrics (pageviews per visit, and revenue per visit), bounce rates, and traffic origin at the dashboard-week level. Dashboards vary in traffic volume from about 7,000 weekly pageviews to almost 5 million weekly pageviews at the 10th and 90th percentiles. The pageview and revenue distributions have long right tails; both means exceed the respective medians by an order of magnitude. This fact motivates our use of logged dependent variables in our analysis. Usage patterns vary by dashboard: on some dashboards users browse less than two pages per visit on average while on others users browse as many as eight. This pattern is evident for the revenue per visit metric as well. Bounce rates are also heterogeneous across sites (10th percentile of 14.5% to 90th percentile of 70.8%) but 40.8% on median. Since traffic from the EU falls under the GDPR's scope,

[10] Specifically, we remove dashboards with average pageviews per visit and average revenue per visit greater than the 95th percentile of the distribution. At the extreme, these outliers are thirty times the mean pageviews per visit and several thousands of times the mean revenue per visit in Table 1. Note that our main effect point estimates in Section 4.2 and our pageview and revenue per visit estimates (in Table 4) are robust to different trimming thresholds, but the latter are imprecise when we include these outliers.

[11] Annual US e-commerce spending in 2017 was $461.5 billion. This is approximately $37.8 billion per month after excluding the large increases in holiday spending in November and December (U.S. Census Bureau, 2019).

[12] https://stats.wikimedia.org/wikimedia/squids/SquidReportPageViewsPerCountryOverview.htm

Table 1: 2018 pre-GDPR weekly summary statistics

Variable	# Dashboards	Mean	10th percentile	Median	90th percentile
Pageviews (full sample)					
EU user pageviews	1,084	4,008,584	6,722	108,349	4,483,751
EU user pageviews per visit	1,084	4.47	1.73	3.64	7.87
EU user bounce rates	1,084	42.3%	14.5%	40.8%	70.8%
% EU/ Global	1,084	45.1%	0.5%	19.3%	99.1%
% North America/ Global	1,084	39.1%	0.4%	10.8%	97.8%
Revenue (e-commerce sample)					
EU user revenue	353	$505,379	$1,507	$47,872	$1,352,089
EU user revenue per visit	353	$3.35	$0.07	$1.63	$9.26
EU user bounce rates	353	37.1%	16.4%	35.3%	61.3%
% EU / Global	353	46.0%	0.1%	21.3%	99.1%
% North America / Global	353	45.4%	0.2%	22.4%	99.6%

Note: The full sample consists of all 1,084 dashboards in our data set. The e-commerce sample consists of the subset of full sample dashboards who report revenue.

we summarize dashboard EU user outcomes as a share of global outcomes (as well as North America's share for comparison). 41.2% of all dashboards receive more than three quarters of their traffic from EU users and 33.7% receive more than three quarters from North America. The sample's traffic tilts towards EU users on average, though the average share of revenue is balanced between EU and North American users. Throughout, the pageviews outcome references the full sample, whereas the revenue outcome references the e-commerce sample.

Our revenue measure captures e-commerce revenue, and excludes other revenue sources like offline revenue and advertising revenue. In Adobe Analytics, firms track revenue at the product level: when the user completes a transaction, the revenue is the product of the product's price and quantity purchased. Elsewhere, this data has been used to track inflation and e-commerce spending (Goolsbee and Klenow, 2018; Adobe, 2021). Three industry categories represent 70% of our revenue data: consumer electronics (34.7%), jewelry & apparel (23.5%), and general merchandise (i.e., department stores, 11.5%). Our revenue data arises from an additional thirteen industries including: airlines, automobiles & parts, books, finance & banking, food & drink, travel, gifts & flowers, health, home & garden, hotels, media & entertainment, shipping, and software. Moreover, our revenue share figures by industry appear to be representative.[13]

Table 2 summarizes the mean last-touch attribution channel shares. Table 2 includes the 542 dashboards (and 260 e-commerce dashboards) that collect this data for at least

[13] https://www.emarketer.com/content/us-ecommerce-by-category-2021

Table 2: Last-touch attribution sample: pre-GDPR mean channel shares

	Recorded pageviews	**Recorded revenue**
Observations	542	260
Direct traffic	33.3%	39.1%
Search	58.1%	51.9%
Display	1.3%	0.4%
Email	3.9%	7.9%
Social	3.4%	0.6%

Note: 2018 EU pre-GDPR outcomes for those dashboards that report at least 3 channels. Channels shares within dashboards are weighted by pageviews or revenue respectively.

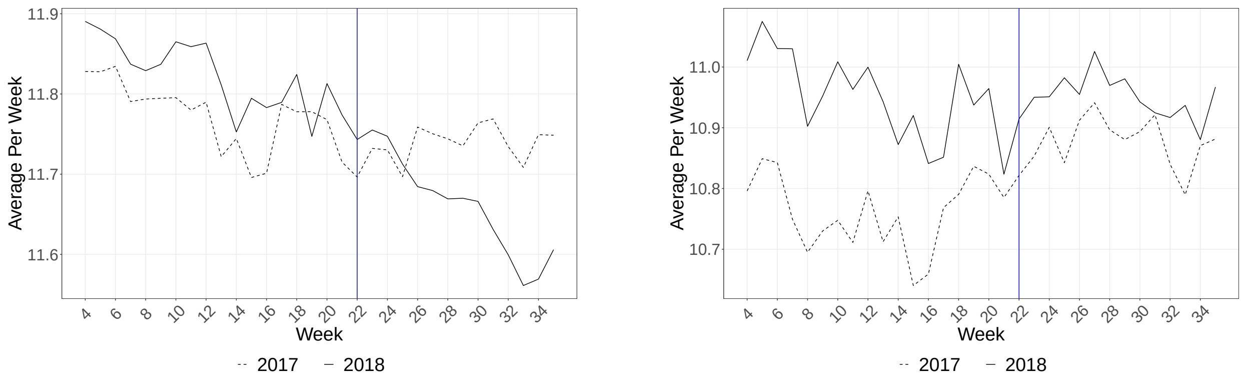
three channels. The last-touch attribution subsample tends to be larger with mean EU weekly pageviews of 8.1 million compared to 4.0 million in the full sample in Table 1. Each dashboard's channel shares are weighted by its pageviews and revenue respectively. The top two channels of direct traffic and search total 91.4% of pageviews and 91.0% of revenue on these sites. Display's referral share is less than 2%, though only 40% of dashboards collect this data for the samples in Table 2. Conditional on reporting, the share of display is 3.3% for all sites and 1.3% for e-commerce. For email, the conditional shares are 5.7% and 7.1% for all sites and e-commerce sites, respectively. By comparison, Budak et al. (2016) report that display ad clicks and email initiate 3% and 7%, respectively, of sessions on the top 10,000 e-commerce sites.

Figure 2 graphs the evolution of log pageviews and of log revenues over time. We use the enforcement of the GDPR—denoted by the vertical line on May 25th, 2018 (week 22)—as an event study in our analysis. Figure 2 also plots same-week 2017 outcomes using a dotted line for comparison. Both pageviews and revenues start off higher in January 2018 than January 2017. The gap is smaller for pageviews, though the two lines cross after the GDPR with 2017 pageviews substantially higher than 2018 by September. Revenue exhibits a larger gap of about 0.2 log points prior to May 25th, but this gap closes substantially afterwards. That is, 2018 revenue is fairly flat, whereas 2017 revenue grows in the second part of 2017. Both outcomes follow similar trends in 2017 and 2018 prior to May 25th, which motivates our choice of 2017 outcomes as a control group in our empirical strategy below.

4 Recorded site outcomes analysis

We seek to quantify the impact of the GDPR on recorded website pageviews and e-commerce revenue. In Section 4.1, we develop our empirical strategy consisting of a panel difference

Figure 2: Evolution of EU user recorded outcomes over time



(a) Log Recorded Pageviews

(b) Log Recorded Revenue

Note: Figure (a) presents average logged pageviews per week for the full sample of dashboards. Figure (b) presents average logged revenue per week for the e-commerce sample of dashboards.

estimator that uses the prior-year's observations for the same firms as a control. Section 4.2 presents the resulting estimates for recorded pageviews and recorded revenue as well as evidence of user behavior changes post-GDPR. Finally, Section 4.3 explores policy-relevant heterogeneity in GDPR treatment effects.

4.1 Panel difference estimator (PD)

The GDPR's scope and imposition creates two challenges for inferring the regulation's effects: 1) identifying an untreated set of dashboards to serve as a control group; and 2) isolating confounding changes post-enforcement. The GDPR's scope extends beyond the EU's borders so that countries outside of the EU may not represent a clean control group. In particular, firms may execute their GDPR compliance strategies across all of their customers in order to reduce customer concerns or the costs associated with treating customers differently by country. Such spillover effects imply that non-EU users are potentially contaminated as a control group. The GDPR's common enforcement date creates a second challenge for inference: post-GDPR changes in outcomes could result from confounding variables that change after May 25, 2018. In Figure 2, we see that pageviews fall post-GDPR, but this may be confounded by seasonal trends like Europe's summer vacation season.

Our control group consists of 2017 outcomes from the same set of dashboards as our treatment group. This ensures that the control group shares seasonal trends and resembles the treatment group. Control units from the past ensures no contamination from the GDPR's enforcement in the EU or its spillover effects beyond the EU. By using the same time period in 2017, this control group accounts for seasonal variation, for instance due to summer vacation and holidays. Further, firms experience different seasonal patterns in demand: surfing websites have more traffic in summer while skiing websites have less. Holding our sample of dashboards constant ensures that the control group captures the seasonal pattern of this specific combination of firms. Our panel differences approach is conceptually similar to interrupted time-series methods (e.g., Baicker and Svoronos, 2019). In interrupted time series, control units are implicitly defined by the continuation of pre-treatment trends. Any difference in trend or level, coinciding with the implementation of treatment, are attributed to the treatment. In our panel differences approach, we allow for a more flexible characterization of control unit behavior by allowing for treatment group-specific seasonality that is not constrained by model functional form.

Figure 2 confirms that pageviews and revenue follow similar trends prior to May 25 in both years. In Online Appendix A.4, we present placebo tests in which we test for pre-treatment differences between our treatment and control group. For both of our outcomes, the results of the placebo treatment exercise add credibility to our panel differences design. The 2017 dashboards are also a natural control group because firm characteristics are relatively fixed year over year. However, we cannot compare the post-May periods between 2018 and 2017 directly because we see in Figure 2 that the pre-May levels exhibit an upward level shift in 2018.

Our primary empirical approach applies a panel difference estimator to compare EU user recorded outcomes in 2018 to 2017. The panel difference estimator is analogous to a difference-in-differences approach in that we assume parallel trends on recorded outcomes as our identifying assumption. That is, the effect of the GDPR on recorded outcomes is identified by the difference in trends post-May between 2017 and 2018. Recall that all real outcomes are recorded pre-GDPR (including 2017), so any post-GDPR deviation from the counterfactual trend must arise from the GDPR's effect on real outcomes and/or its consent effect on data recording (see equation 2). Our regression equation is as follows:

$$\log\left(y_{itw}+1\right)=\alpha\mathbb{1}\{2018\}_t+\beta\left(\mathbb{1}\{2018\}_t \text{ x } \mathbb{1}\{\text{Post-GDPR}\}_w\right)+\eta_i+\xi_w+\epsilon_{itw} \tag{3}$$

where i denotes the analytics dashboard, w denotes calendar week, and t denotes year. y_{itw} denotes *recorded* onsite outcomes by EU users. Our model uses log outcomes because the distributions of recorded outcomes are highly skewed (see Table 1).[14] $\mathbb{1}\{2018\}$ is an indicator variable for 2018 observations (treatment) and $\mathbb{1}\{\text{Post-GDPR}\}$ is an indicator variable for

[14] Our results are robust to other transformations of the dependent variable including: (1) $\log\left(y_{itw}+0.1\right)$, (2) $\log\left(y_{itw}+10\right)$, and (3) an assignment rule that sets all instances of $y_{itw}=0$ to the 1st percentile of y. We estimate a GDPR effect of -0.124 (0.026), -0.121 (0.026), and -0.123 (0.026), respectively for pageviews and -0.143 (0.030), -0.139 (0.029) and -0.142 (0.030), for revenue.

calendar weeks after the May 25th enforcement date. We include dashboard- and calendar-week-specific fixed effects denoted by η_i and ξ_ω. Note that we omit the $\mathbb{1}\{\text{Post-GDPR}\}$ term as it is collinear with the calendar-week fixed effects. β represents the estimate of the average effect of the GDPR on EU user recorded site outcomes in our sample.

4.2 GDPR main effect estimates

Table 3 presents the results of our main specification in equation (3). We estimate main effect coefficients of -0.124 for recorded pageviews and -0.142 for recorded revenues. Both estimates are significant at the 1 percent level.[15] To aid interpretation of our non-linear model, we calculate marginal effects (see Online Appendix C for details). Our marginal-effect estimates indicate a 11.7% drop in recorded pageviews and a 13.3% drop in recorded revenue. For the median dashboard in each sample, this corresponds to a 15,043 drop in weekly recorded pageviews and a $9,227 drop in weekly recorded revenue, respectively.

In Online Appendix A, we provide several robustness checks that allow for alternative control groups and specifications. In particular, we include (1) a difference-in-differences specification with North America as a control group (2) a triple difference-in-differences strategy combining both the panel and North American difference-in-differences specifications (3) window regressions that attempt to remove anticipatory or delayed compliance and (4) a synthetic control approach. The GDPR had substantial spillovers to non-EU users as many websites implemented their compliance approaches globally (Peukert et al., 2022), implying that the GDPR may have lead to drops in recorded outcomes for North American traffic as well as EU traffic. The results of our first two robustness specifications are consistent with this observation: using North America as a control yields attenuated (or equivalent) GDPR treatment effect estimates. The window regressions instead yield a somewhat larger reduction in our estimates. However, all robustness exercise estimates remain statistically significant with the exception of the synthetic control estimate for revenue—which is also smaller in magnitude (-1.4%)—suggesting the revenue results may be somewhat less robust than the pageviews results.

The resulting policy implications vary by whether these estimates reflect changes in real web outcomes, or changes in the share of recorded outcomes. At one extreme, a 12% reduction in real web outcomes would represent a significant economic burden of the GDPR. At the other extreme, a 12% non-consent rate may be construed as progress toward protecting

[15] For robustness, we also employ the permutation inference method outlined in Conley and Taber (2011). Paz and West (2019) suggest that this method of inference is often more conservative than clustered standard errors. Our main effect estimates for pageviews and revenue remain significant at the 10 percent level. Also, we estimate a similar -0.118 main effect coefficient on pageviews for the e-commerce sample (see Online Appendix A.1).

Table 3: GDPR main effect estimates: panel difference estimator

	(1)	(2)
Sample	All Dashboards	E-commerce
Dependent variable†	log(Pageviews + 1)	log(Revenue + 1)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post-GDPR}\}$	-0.124	-0.142
	(0.026)	(0.030)
$\mathbb{1}\{2018\}$	0.050	0.193
	(0.014)	(0.028)
Average marginal effect	-11.67%	-13.26%
RSID + Week FE	Y	Y
R^2	0.969	0.962
N	69,344	22,436

Note: Standard errors clustered at the Dashboard + Week level. †Recorded outcomes. See Online Appendix A.1 (Table 9) for the e-commerce sample's pageviews estimate.

the data of privacy-sensitive users. The main effect estimates should reflect something between these two extremes: a mixture of a real and recording effect. Section 5 proceeds to construct bounds on how much the GDPR affects real outcomes versus consent by using a theory-driven empirical analysis.

We next consider how the above changes relate to the change in site visits (extensive margin) and the pageview/revenue per visit (intensive margin). The types of user who use a site and provide consent may change in response to the GDPR. While we do not observe individual user behavior in our data, we do observe two metrics that inform how users behave on-site: pageviews per visit, and revenue per visit. Collectively, we refer to these outcomes as *site-usage* metrics. Table 4 presents our GDPR impact estimates employing the panel differences specification (equation 3) for the site-usage metrics (in levels) as dependent variables. Table 4 also presents comparable estimates for the extensive margin of site visits (in logs). Note that our mechanism bounding approach (Section 5) uses both the intensive and extensive margins.

Our usage estimates show that recorded users are favorably selected post-GDPR. In columns (1) and (2) of Table 4, we see a statistically significant increase in pageviews per visit of 0.200 (s.e. 0.071) and a marginally significant increase in revenue per visit of \$0.172 (\$0.093). Our average dashboard sees 4.47 pageviews per visit in the pre-GDPR period, so that our estimate represents a 4.5% increase in pageviews per visit post-GDPR. For e-commerce dashboards, our point estimate is noisier but corresponds to a 5.1% increase in revenue per visit. The extensive margin estimates in columns (3) and (4) of Table 4 yield almost identical estimates to Table 3: a -0.128 reduction in log visits for the full sample and -0.148 for the e-commerce sample.

Table 4: GDPR effect estimates: intensive and extensive margins

	Usage metrics		*Extensive Margin*	
	(1)	(2)	(3)	(4)
Sample	All Dashboards	E-commerce	All Dashboards	E-commerce
Dependent variable†	Pageviews per visit	Revenue per visit	log(Visits + 1)	log(Visits + 1)
$\mathbb{1}\{2018\}\times$ $\mathbb{1}\{\text{Post GDPR}\}$	0.200	0.172	-0.128	-0.148
	(0.071)	(0.093)	(0.024)	(0.031)
$\mathbb{1}\{2018\}$	-0.291	0.041	0.102	0.167
	(0.067)	(0.094)	(0.012)	(0.022)
RSID + Week FE	Y	Y	Y	Y
R^2	0.724	0.758	0.971	0.961
N	69,344	22,436	69,344	22,436

Note: Standard errors clustered at the Dashboard + Week level. †Recorded outcomes.

We argue that the increase in recorded usage is primarily due to consent-based selection: that is, users who consent to data processing have greater usage. This could arise if users who value the site more are more likely to both provide consent and to browse/spend more onsite. We argue that consent-based selection is a more plausible explanation for the observed positive selection than the GDPR increasing real site usage. If anything, we instead expect that the GDPR would hurt real usage by degrading the browsing experience with cookie consent and notification pop-ups. In particular, we expected that these popups would increase the share of visits that include only a single page (i.e., the bounce rate), however we find no empirical support for this (see Online Appendix D for details). Note that Aridor et al. (2020) also show favorable selection in usage outcomes post-GDPR for online travel agencies, which they also argue is attributable to user consent. We return to the observation of favorable user selection in Section 5, where we use this pattern to decompose our main effect estimate into real and consent effects.

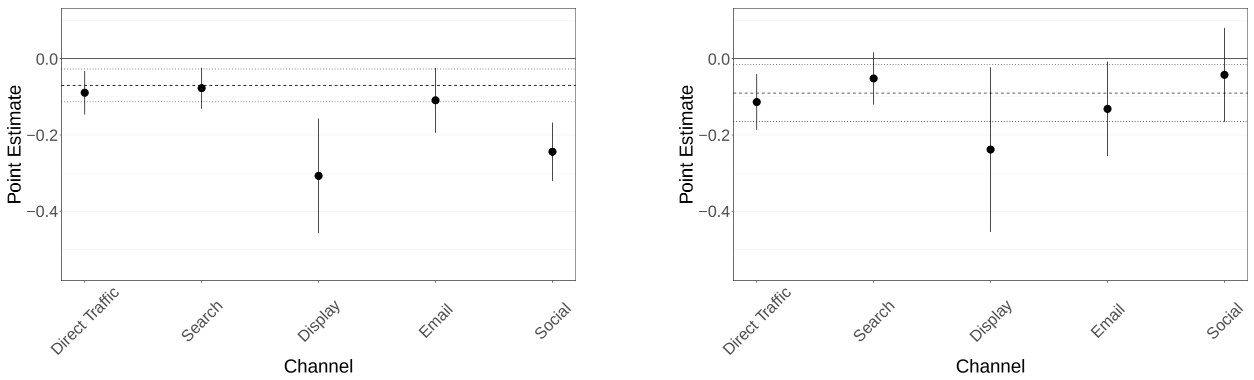
4.3 GDPR effect heterogeneity

To better understand the GDPR's impact, we examine heterogeneity in terms of traffic referral channels, dashboard size, and regulatory strictness.

4.3.1 Traffic referral channels

As in Section 2, we expect that the GDPR may harm websites by constraining their marketing activities. In particular, by increasing the costs of processing personal data, the GDPR may have limited the use and effectiveness of personalized marketing channels like email and display advertising. To examine this, we consider heterogeneity in the impact of the GDPR

Figure 3: GDPR effect heterogeneity by traffic referral channel: last-touch attribution sample



(a) Pageviews

(b) Revenue

Note: Presents the channel-level coefficients (equation 27) with recorded dependent variables in logs and their associated 95% confidence intervals using standard errors clustered at the dashboard-week level. The dashed line plots the overall mean effect of the GDPR across all channels, with dotted lines indicating the corresponding 95% confidence interval.

by traffic referral channel using the "last-touch attribution" sample of 542 dashboards that track this data for at least three channels. Specifically, these dashboards track whether users click on a search, display, social, or email advertisement to arrive on site—or instead navigate directly to the site. Personalized marketing is an important source of traffic for these dashboards: Table 2 shows that e-mail clicks precede 3.9% of pageviews and 7.9% of e-commerce revenue while display ad clicks precede 1.3% of pageviews and 0.4% of e-commerce revenue.

We note that the dashboards that track traffic referral channel data are selected. These firms may be more sophisticated marketers or may rely more on marketing to generate site traffic. Though the 542 dashboards in our last-touch attribution sample are larger (see Section 3.4), they exhibit somewhat smaller GDPR main effect estimates. For this sample, we estimate a GDPR effect of -0.07 (s.e. 0.022) log points for pageviews with a corresponding marginal effect of -7.1% and -0.09 (s.e. 0.038) log points for revenue with a marginal effect of -8.7%. These estimates are smaller than their full sample counterparts in Table 3, but this difference is only statistically significant for pageviews and not revenue.

Figure 3 suggests post-GDPR declines in all channels for both pageviews and e-commerce revenue,[16] and particularly large declines in the email and display ad channels that process personal data. Figure 3 presents point estimates, though we refer to the corresponding marginal effects below. The direct traffic channel serves as a useful benchmark; these esti-

[16] We present the exact estimation equation and coefficient estimates in Online Appendix 3. Note that Figure 3's channel estimates each condition on the set of dashboards that collect data for that channel. Variation in sample size by channel thus explains the large differences in confidence interval widths.

mates suggest statistically significant drops of -8.5% for pageviews and -10.7% for revenue. Personal marketing referral channels show both statistically significant and larger reductions. The display ad channel in particular falls -26.5% for pageviews and 21.2% for revenue whereas the email channel falls -10.3% for pageviews and -12.3% for revenue.

The evidence for the search and social channels is more mixed. The search channel accounts for over half the pageviews and e-commerce revenue in the last-touch attribution sample (Table 2). The search channel falls by less than direct traffic for both pageviews (-7.4%) and e-commerce revenue (-5.0%), though only the former is statistically significantly different from zero. Several factors may offset the negative effect of consent leading to smaller effects on the search channel. Search is less susceptible to the GDPR because it does not require personal data, which may have led firms to reallocate their marketing budgets towards the paid search channel. Also, users may have increased their search activity post-GDPR (Zhao et al., 2021). The social channel represents a much smaller share of traffic for these dashboards. Relative to direct traffic, we see a larger drop in the social advertising channel for pageviews (-21.4%), but a smaller one for e-commerce revenue (-3.4%), though again only the former is statistically significantly different from zero. The social ad channel occupies a more nebulous standing with the GDPR because it relies on personal data, though much of it is first-party data. In practice, social media companies appeared to fare well at least in the first year of the GDPR. For e-commerce firms, the smaller reduction in the social channel may reflect these firms reallocating their marketing budgets, as with the search channel.

In sum, we find greater reductions in recorded pageviews and revenue post-GDPR in the personalized marketing channels targeted by the regulation. We return to these results in Section 5.3, where we construct conservative lower bounds on the personalized marketing component of the GDPR's real effect on websites.

4.3.2 Dashboard size

Here, we examine heterogeneity in GDPR treatment effect estimates as a function of dashboard size. The competitive ramifications of the GDPR are a key point of policy interest. Other work demonstrates that the GDPR led to greater market concentration in the website technology services sector (Johnson et al., 2022; Peukert et al., 2022) and has more adversely impacted investment in newer technology firms (Jia et al., 2021). Larger sites may benefit from the GDPR as they may more easily obtain consent (Campbell et al., 2015) or may be less reliant on push marketing.

We look for heterogeneous effects of the GDPR across dashboard size by interacting our main specification (equation 3) with an indicator for large dashboards. We define large dashboards as having pre-GDPR average weekly visits greater than the median of the distribution.

Table 5: GDPR effect heterogeneity by dashboard size

	(1)	(2)
Dependent Variable†	log(Pageviews + 1)	log(Revenue + 1)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \mathbb{1}\{\text{Large Dashboard}\}$	0.040	0.101
	(0.034)	(0.053)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\}$	-0.144	-0.182
	(0.034)	(0.046)
Average Marginal Effect		
Small Dashboard	-13.43%	-16.69%
Large Dashboard	-9.87%	-7.85%

Note: See Table 14 in Online Appendix B.1.3 for full table. Standard errors clustered at the Dashboard + Week level. †Recorded outcomes.

Table 5 presents the key parameter estimates: Online Appendix B.1.2 presents the full estimation equation and results. The base coefficient ($\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\}$) estimates indicate that the GDPR had a negative and significant impact on small dashboards for both recorded pageviews and revenue. The large dashboards interaction coefficient estimates are positive for both recorded pageviews and revenue. This indicates that large dashboards experience a smaller post-GDPR declines than small dashboards, though this difference is only statistically significant for the revenue outcome. This pattern can be seen in the marginal effect estimates in Table 5: small dashboards see a -13.4% decline in pageview (versus -9.9% for large) and a -16.7% decline for revenue (versus -7.9% for large). In Section 5.2.2, we provide evidence that heterogeneous consent rates help explain this result.

4.3.3 Regulatory strictness

We examine whether firm beliefs about regulatory strictness moderate the effect of the GDPR. Firms report heterogeneous investment in and readiness for the GDPR (TrustArc, 2018). Regulators can vary the size of the fines and the probability of levying fines to induce firms to comply with regulation. As such, firms that face a more strict regulator may work harder to comply with the law. For instance, these firms may cut back on personalized marketing or set a higher bar for collecting consent on their website. This exercise is interesting from a regulatory design perspective and provides additional evidence that the GDPR explains the changes in EU traffic after May 25, 2018.

To examine the role of regulatory strictness, we exploit geographic variation in firm's beliefs about their country's data protection authority. Though the GDPR harmonizes privacy regulation in the EU, the GDPR's enforcement is handled by authorities in each EU country. Firms are assigned to a country's data protection authority based on where their

Table 6: GDPR effect heterogeneity by regulatory strictness

	(1)	(2)
Dependent Variable†	log(Pageviews + 1)	log(Revenue + 1)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \text{Strictness}$	-0.040	-0.041
	(0.006)	(0.019)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\}$	0.041	-0.040
	(0.029)	(0.075)

Note: See Table 13 in Online Appendix B.1.2 for full table. The strictness index is normalized to have a mean of zero and standard deviation of one. Standard errors clustered at the Dashboard + Week level. †Recorded outcomes.

customers or headquarters are located. To quantify regulatory strictness by EU country, we use of a measure from the European Commission (2008) survey of 4,835 data controllers. Controllers report whether data protection law is "interpreted and applied more rigorously" in their country than in the rest of the EU. We construct an mean zero index that ranges from -1.64 in Greece to 1.49 in Sweden.[17]We assign each dashboard to an EU country based on the member state that drives the majority of its web traffic in 2018 before the GDPR.

Our resulting estimates reject the null hypothesis that regulatory strictness is unrelated to the GDPR's effect on recorded web outcomes. We re-estimate our panel differences model (equation 3), further including an interaction with regulatory strictness. We also include another interaction term for income per capita, which may confound the role of regulatory strictness. We present the key estimates in Table 6, and the full regression results and model specification in Online Appendix B.1.3. We estimate a regulatory strictness interaction coefficient of -0.040 log points for pageviews and -0.041 for revenue, which are both significant at the 1% level. The estimates imply that a one standard deviation increase in regulatory strictness, holding all else fixed, reduces recorded pageviews by -4.0% and recorded revenue by -4.1%.

Thus, country-level differences persist in practice, despite the GDPR's promise of a consistent regulatory environment within the EU. We acknowledge that this strictness index may proxy for other cross-country differences like norm adherence or bureaucratic efficiency. Nevertheless, the moderating role of this regulatory strictness measure is a robust finding of the GDPR literature. Johnson et al. (2022) also show that regulatory strictness is correlated with website changes to their tech vendor use post-GDPR. Jia et al. (2020, 2021) show that regulatory strictness is correlated with post-GDPR reductions in EU tech venture investment. These findings further substantiate that our main effect estimates capture the

[17] After excluding non-responses, the survey responses takes four values from "totally disagree" (coded as 1) to "totally agree" (coded as 4). We average survey responses by country and then normalize the country-level index to be mean zero and standard deviation of one.

impact of the GDPR rather than an unrelated post-enforcement shock to EU web traffic.

5 Disentangling real and recorded outcomes

To better evaluate the impact of the GDPR on the online economy, we need to separate the measured effect on recorded outcomes from its effect on real outcomes (equation 2). Separating real and recording effects in our setting is difficult for two reasons. First, we do not observe user-level behavior. Rather, we observe summary statistics that conflate recording and real outcomes at the dashboard level. Second, we do not directly observe the compliance strategies of each dashboard and thus cannot observe how recording and consent rates vary with individual dashboard compliance.

In the next section, we develop a model of the underlying data generating process that highlights how, and under what assumptions, we can separately recover estimates of the GDPR's real and recording effects.

5.1 A model of recorded outcomes

Let $Z \in \{0,1\}$ be an indicator for the GDPR treatment. For user interactions with a representative website, we define three potential outcomes. Let $V\left(Z\right) \in \{0,1\}$ be an indicator for whether a user decides to visit a website, let $C\left(Z=1\right) \in \{0,1\}$ be an indicator for whether the user consents to sharing data with the website under the GDPR treatment, and let $y\left(Z\right)$ denote the user's (continuous) usage of the website as measured in pageviews or revenue. Note that the consent potential outcome (C) is only realized with the GDPR treatment ($Z = 1$) whereas the visit (V) and usage (y) potential outcomes can have different realizations without and with GDPR treatment ($Z \in \{0,1\}$).

We model *real outcomes* as realizations of the potential outcomes $\{V(Z), C(Z=1), y(Z)\}$. Reflecting the GDPR's censoring of non-consenting user data, we then define individual-level *recorded outcomes* as follows:

$$V^{obs}(Z) = \begin{cases} V\left(0\right) & if\, Z=0 \\ V\left(1\right) \cdot C\left(1\right) & if\, Z=1 \end{cases} \tag{4}$$

$$y^{obs}(Z) = \begin{cases} V\left(0\right) \cdot y\left(0\right) & if\, Z=0 \\ V\left(1\right) \cdot C\left(1\right) \cdot y\left(1\right) & if\, Z=1 \end{cases} \tag{5}$$

Though we do not directly observe individual level outcomes, V^{obs} and y^{obs}, we can infer their conditional means from two observe quantities at the site level. Defining the unobserved number of total possible users as M, we can write our observed site-level outcomes—i.e., visits and usage—as functions of the unobserved individual outcomes defined above:

$$N^{obs}(Z) = \begin{cases} M \cdot E\left[V\left(0\right)\right] & if\, Z = 0 \\ M \cdot E\left[V\left(1\right) \cdot C\left(1\right)\right] & if\, Z = 1 \end{cases} \tag{6}$$

$$E\left[y^{obs}(Z) \,|\, V^{obs}(Z) = 1\right] = \begin{cases} E\left[y\left(0\right) \,|\, V\left(0\right) = 1\right] & if\, Z = 0 \\ E\left[y\left(1\right) \,|\, V\left(1\right) = 1, C\left(1\right) = 1\right] & if\, Z = 1 \end{cases} \tag{7}$$

Equation (6) captures the *total recorded visits* and equation (7) captures the *mean recorded usage*, where expectations are taken across individual users.

Without further assumptions, inference is intractable. Our identification approach classifies users into types as defined by their (binary) potential outcomes $V(0)$, $V(1)$, and $C(1)$. Table 7 enumerates all eight possible types, corresponding to {0,1} outcomes for site visits with/without the GDPR and consent with the GDPR. However, types for which $V(0) = V(1) = 0$ are immaterial to our analysis as such users do not visit the site regardless of the GDPR. To make progress, we posit plausible restrictions on the composition of the six remaining user types, along with related assumptions on usage conditional upon user type:

Assumptions

1. The GDPR does not induce new visits: $V\left(0\right) = 0 \Longrightarrow V\left(1\right) = 0$. In particular, this rules out user types for which $V(0) = 0$ and $V(1) = 1$.

2. Usage is conditionally independent of the GDPR, given a user's visit and GDPR consent type: $(y \perp Z) \,|\, V\left(0\right), V\left(1\right), C\left(1\right)$

3. The user's choice of consent is independent of their choice to visit: $C\left(1\right) \perp V\left(1\right)$

4. Only consent-based variation in average usage:
$E\left[y|V\left(0\right) = 1, C\left(1\right)\right] = E\left[y|V\left(1\right) = 1, C\left(1\right)\right] = E\left[y|C\left(1\right)\right]$

Assumption 1 rules out user types that visit under the GDPR but would not visit in the absence of the GDPR (regardless of consent). Our identification approach thus classifies users into four remaining types, defined by their counterfactual behavior under the GDPR: 1) users who visit both before and after the GDPR and consent; 2) users who visit both before and after the GDPR, but do not consent; 3) users who visit pre-GDPR, do not visit post-GDPR, but would have consented if they had visited; and 4) users who visit pre-GDPR, do not visit post-GDPR, and would not have consented if they had visited.

Table 7: User types defined by discrete potential outcomes

Type	V(0)	V(1)	C(1)	Comments
1	1	1	1	allowed: same expected usage as type 3 (Assumption 4)
2	1	1	0	allowed: same expected usage as type 4 (Assumption 4)
3	1	0	1	allowed: same expected usage as type 1 (Assumption 4)
4	1	0	0	allowed: same expected usage as type 2 (Assumption 4)
5	0	1	1	ruled out by Assumption 1
6	0	1	0	ruled out by Assumption 1
7	0	0	1	immaterial to analysis
8	0	0	0	immaterial to analysis

Note: {V(0), V(1)} = visit {without, with} GDPR, C(1) = consent with GDPR

Assumption 2 requires that the user's browsing and spending choices are unaffected by the GDPR, conditional on the user's visit and consent choices. Assumptions 1 and 2 together imply that the GDPR does not directly impact site usage: rather, it operates indirectly through consent-based selection. A potential concern with Assumption 2 is that it rules out GDPR-related browsing frictions conditional on visiting a site. However, this assumption is consistent with our lack of evidence supporting this type of privacy friction (see Online Appendix D). The independence of visit and consent decisions (Assumption 3) is plausible to the extent that users are myopic or do not have well-formed expectations about how a site will comply with GDPR. Finally, Assumption 4 restricts site usage to be the same (on average) among users of the same consent type, regardless of their choice to visit the site under GDPR. Assumption 4, while strong, helps us to determine the relative size of the four user types by imposing that types one and three, as well as types two and four, have the same expected usage.

Consistent with the above assumptions, we define two quantities of interest at the visit level, which correspond to the real and consent effects of the GDPR. We define the non-consent rate $\theta \in [0,1]$, and the real effect of the GDPR $\delta \in [0,1]$, as:

$$\theta \equiv 1 - E\left[C\left(1\right)\right] \tag{8}$$

$$\delta \equiv 1 - \frac{E\left[V\left(1\right)\right]}{E\left[V\left(0\right)\right]} \tag{9}$$

Note that, in equations (8) and (9), θ and δ are defined in terms of potential outcomes.

Proposition 1 Under Assumptions 1-4, θ and δ are set identified, as follows:

A) the non-consent rate (θ) is given by:

$$\theta = \frac{E\left[y^{obs}|V^{obs}=1, Z=1\right] - E\left[y^{obs}|V^{obs}=1, Z=0\right]}{E\left[y^{obs}|V^{obs}=1, Z=1\right] - E\left[y|V\left(0\right)=1, C\left(1\right)=0\right]} \tag{10}$$

B) the decline in recorded visits is a product of the consent (θ) and real (δ) effects of the GDPR:

$$\frac{E\left[V^{obs}=1|Z=1\right]}{E\left[V^{obs}=1|Z=0\right]} = \left(1-\delta\right)\left(1-\theta\right) \tag{11}$$

We sketch the proof below, and provide the details in Online Appendix G. For part (A), we begin by noting that the law of total expectation and Assumption 3 imply we can calculate expected usage without the GDPR as a consent probability-weighted average:

$$\begin{aligned} E\left[y|V\left(0\right)=1\right] = & \; E\left[y|V\left(0\right)=1, C\left(1\right)=0\right]\cdot\theta \\ & + E\left[y|V\left(0\right)=1, C\left(1\right)=1\right]\cdot\left(1-\theta\right) \end{aligned}$$

Next we note that by Assumption 2, the left-hand side expression is equivalent to observed usage conditional on visiting without the GDPR, $E\left[y|V\left(0\right)=1\right] = E\left[y^{obs}|V^{obs}=1, Z=0\right]$. By Assumption 4, we can also replace the second expectation in the right-hand side expression with an observed quantity, $E\left[y|V\left(0\right)=1, C\left(1\right)=1\right] = E\left[y^{obs}|V^{obs}=1, Z=1\right]$. Substituting and solving for θ yields equation (10). Finally, part (B) follows from Assumption 3 and our definitions of δ and θ.

Proposition 1 part (A) shows that the non-consent rate θ depends on the unobserved pre-GDPR usage among users who would not provide consent: $E\left[y|V\left(0\right)=1, C\left(1\right)=0\right]$. Below, we show that we can partially identify θ using bounds on $E\left[y|V\left(1\right)=1, C\left(1\right)=0\right]$. Further, note that equation (10) is only informative if usage differs between users who do and do not consent ($E\left[y^{obs}|V^{obs}=1, Z=1\right] \neq E\left[y|V\left(1\right)=1, C\left(1\right)=0\right]$). The results in Table 4 show that recorded usage increases post-GDPR, which suggests that users who consent are indeed favorably selected.

Proposition 1 part (B) relates the ratio of observed visits with and without the GDPR to the real and consent effects. Assumption 3 plays a critical role in the derivation of equation (11), as it allows both θ and δ to enter multiplicatively and independently on the right-hand side. Through equation (11), the θ bounds imply corresponding bounds for the real effect of the GDPR, δ. We use under- and over-bars to denote lower and upper bounds, respectively.

Corollary 1 $\left(\underline{\theta}, \bar{\delta}\right)$ A lower bound on the non-consent rate and an upper bound on the real effect are given by:

$$\underline{\theta} = \frac{E\left[y^{obs}|V^{obs}=1,Z=1\right] - E\left[y^{obs}|V^{obs}=1,Z=0\right]}{E\left[y^{obs}|V^{obs}=1,Z=1\right] - \underline{y}} \tag{12}$$

$$\bar{\delta} = 1 - \frac{E\left[V^{obs}|Z=1\right]}{E\left[V^{obs}|Z=0\right]\cdot\left(1-\underline{\theta}\right)} \tag{13}$$

Corollary 1's $\underline{\theta}$ follows from equation (10). Note that equation (10) implies an increasing relationship between θ and the unobservable $E\left[y|V(1)=1,C(1)=0\right]$. However, this unobservable has the lower bound $E\left[y|V(1)=1,C(1)=0\right] \geq \underline{y}$. Note that $\underline{y}=1$ for the pageviews-per-visit outcome because each visit mechanically has at least one pageview and $\underline{y}=0$ for the revenue-per-visit outcome. The upper bound on $\bar{\delta}$ is then given by substituting $\underline{\theta}$ into equation (11).

Corollary 2 $\left(\bar{\theta},\underline{\delta}\right)$ An upper bound on the non-consent rate and a lower bound on the real effect are given by:

$$\bar{\theta} = 1 - \frac{E\left[V^{obs}|Z=1\right]}{E\left[V^{obs}|Z=0\right]} \tag{14}$$

$$\underline{\delta} = 0 \tag{15}$$

Intuitively, the upper bound $\bar{\theta}$ (and lower bound $\underline{\delta}$) occurs when the real effect is minimized. Note that Online Appendix E includes a numerical example and graphical proof of our bounding analysis to provide further intuition.

5.2 Estimation & Results

We next provide estimates for the parameter bounds $\left[\underline{\delta},\bar{\delta}\right]$ and $\left[\underline{\theta},\bar{\theta}\right]$. To do so, we use the relations in Corollaries 1 and 2 to define site-level moment equations and estimate these parameters using GMM. We provide detailed estimation equations in Online Appendix F. Our estimation approach integrates our panel differences framework, which accommodates seasonal trends.

5.2.1 Main effect bounds

Figure 4 reports our bounds estimates first in terms of visits (Figure 4a) and then in terms of pageviews/revenue (Figure 4b). In the latter case, the lower bound consent estimates are rescaled by worst-case usage rates (i.e. $\underline{\theta}\cdot\underline{y}$) leading to smaller estimates. Figure 4a presents the bounds estimates for the full and e-commerce samples as a percentage of pre-GDPR visits. For the full sample, the non-consent rate (θ) falls between [4.0%, 12.8%]. For the e-commerce sample, our θ bound estimates are somewhat wider: [3.6%, 13.7%]. These estimates are consistent with contemporary industry reports indicating non-consent rates less than 10% (Quantcast, 2018). These consent rates are high and consistent with the opt-out consent standard that emerged post-GDPR despite the regulation's opt-in consent requirement.

For greater economic interpretability, we convert these visit-level consent bounds into pageview and revenue terms presented in Figure 4b. For all sites, the pageview-level non-consent rate bounds are [0.6%, 11.7%]. The pageviews upper bound is given by the GDPR average marginal effect estimate on recorded pageviews from Table 3. The lower bound of 0.6% is conservative because we assume pageviews per visit for non-consenting users is equal to one. The real effect bounds are then [0%, 8.1%]. However, the equivalent revenue-level non-consent and real effect bounds for the e-commerce sample are uninformative: [0%, 13.2%].[18]

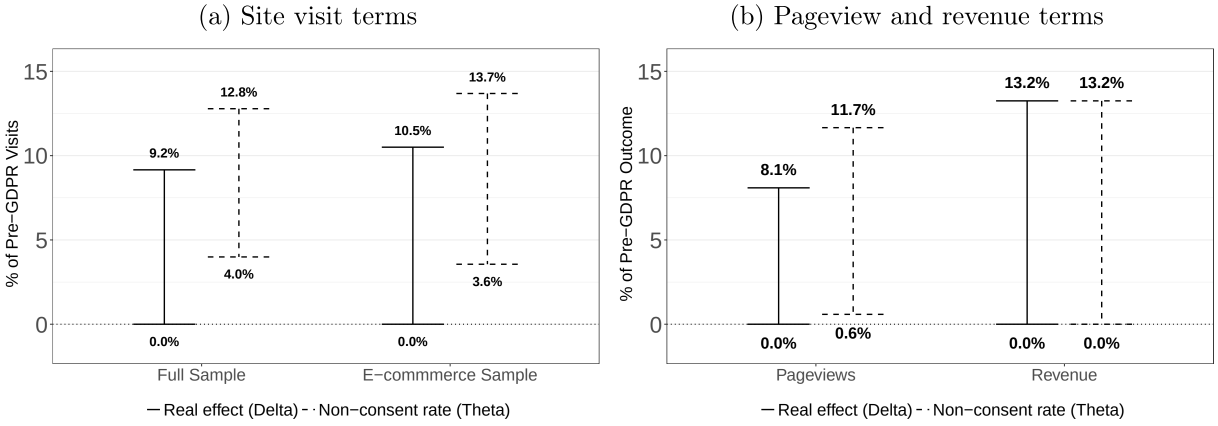
Using the informative pageview bounds in Figure 4b, we can further quantify the minimum contribution of non-consent to the average marginal effect of GDPR on observed pageviews from Table 3 (-11.7%). For this, we conservatively assume a minimal consent effect ($\underline{\theta}$) combined with a maximal real effect ($\bar{\delta}$), which then implies that consent accounts for at least 4.7% of our overall GDPR average marginal effect estimate on pageviews.[19] In place of an equivalent revenue-level figure, we note that the lower bound on visits ($\underline{\theta}$) for e-commerce sites is 26% of the upper bound ($\bar{\theta}$) that assumes all of the GDPR effect is due to consent. We conclude that the effect of the GDPR on data recording—via user consent—contributes materially to our GDPR estimates on recorded site outcomes.

We note too that consent rates can have real consequences for firms. First, online firms can use analytics data to improve business outcomes (Berman and Israeli, 2022), but firms will have both less data and selected data post-GDPR. Second, many websites rely on revenue from personalized advertising that relies on third-party cookies. Two recent studies on the value of personalized advertising both found that ad prices fall by 52% without personalization (Johnson et al., 2020; Ravichandran and Korula, 2019), while the Competition and Markets Authority (2020) report suggests this figure is conservative. Combining the 52%

[18] The upper bound is the recorded revenue estimate from Table 3. However, the lower bound in revenue terms is uninformative because it conservatively assumes the lower bound $E\left[y|V\left(1\right)=1, C\left(1\right)=0\right]=\$0$; in other words, the non-consenting users have no revenue impact because they are assumed to generate no revenue at this lower bound.

[19] In order to account for the real effect on arrivals, we first scale the consent effect down to the population of arriving users and then divide by the main effect estimate: $\frac{\underline{\theta}\cdot\left(1-\bar{\delta}\right)}{\beta}=\frac{0.006\cdot(1-0.081)}{0.117}$.

Figure 4: GDPR mechanism bound estimates



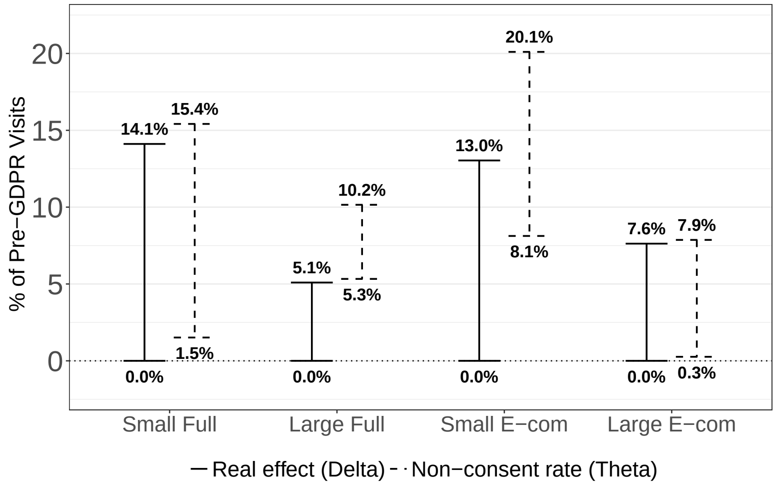
Note: Figure A presents the results of the bounding analysis for both samples in terms of pre-GDPR visits. Figure B presents an analogous figure in terms of (1) pageviews for the full sample and (2) revenue for the e-commerce sample.

figure with our pageview-level non-consent bound estimates, we obtain that websites with advertising would lose between 0.3% and 6.1% in ad revenue due to non-consenting users.

5.2.2 Dashboard size heterogeneity

Further pursuing our Section 4.3.2 result that smaller dashboards exhibit greater GDPR main effect estimates, Figure 5 below presents the real and consent effect bounds for both samples, split by dashboard size. In the full sample, we estimate non-consent rates between [1.5%, 15.4%] for small dashboards and [5.3%, 10.2%] for large dashboards. These consent bounds imply an upper bound on the real effect of 14.1% for small dashboards and 5.1% for large dashboards. These bounds overlap, making comparison difficult, but may suggest that larger firms see smaller real effects post-GDPR. The results for the e-commerce sample are more stark. We estimate non-consent rates between [8.1%, 20.1%] for small dashboards and [0.3%, 7.9%] for large dashboards. These regions do not overlap, suggesting that large dashboards, at least in the e-commerce sample, are more likely to receive consent from users post-GDPR. These results support the argument made by Campbell et al. (2015): larger firms may be able to gain consumer consent more easily. Further, this advantage for larger firms may compound in the long run if firms can use consumer data to improve their decision-making.

Figure 5: Heterogeneity in mechanism bounds by dashboard size



Note: This figure breaks the bound analysis out by site size using a median split. All results are presented in terms of pre-GDPR visits.

5.3 Refining the real effect lower bound

The bounds estimates from Section 5.2 are uninformative of the lower bound on the real effect of the GDPR ($\underline{\delta}$). Nevertheless, the point estimates in Section 4.3.1 suggest a specific explanation for a real effect of the GDPR: i.e., personalized marketing may have been adversely affected by the GDPR. To the extent that marketing is harmed by the GDPR, it contributes to the real effect (δ) as defined in equation (9). In this section, we use heterogeneity in GDPR treatment effects across referral channels to refine our estimate of the lower-bound real effect of the GDPR. To accomplish this, we extend our model from Section 5.1 to accommodate multiple channels and apply this logic to the subset of dashboards that report channel referral data. We then employ an additional assumption to isolate a conservative estimate of the GDPR's marketing effect on consenting users, and thereby refine our real effect lower bound. To begin, we extend the model to accommodate multiple personalized marketing channels of arrival, A, and make the following assumption in addition to those made in Section 5.1:

Assumption 5 The direct traffic channel places an upper bound on the non-consent rate for the personalized marketing channels: i.e., $\theta_{\text{Direct Traffic}} \geq \theta_{\text{A}}$ for $A \in \{\text{Display, Email}\}$.

Assumption 5 pins down an upper bound for the non-consent rate using the *direct traffic* attribution channel. Assumption 5 is consistent with Figure 3 in that the reduction in visits post-GDPR is smallest for direct traffic.

Proposition 2 Under assumptions 1-5,

A) An upper bound for the consent effect on direct traffic is given by:

$$\bar{\theta}_{\text{Direct Traffic}} = 1 - \frac{E\left[V^{obs}_{\text{Direct Traffic}} | Z = 1\right]}{E\left[V^{obs}_{\text{Direct Traffic}} | Z = 0\right]} \tag{16}$$

B) Lower bounds on the real effect in the display and email channels are given by:

$$\underline{\delta}_{\text{Display}} = 1 - \frac{E\left[V^{obs}_{\text{Display}} = 1 | Z = 1\right]}{E\left[V^{obs}_{\text{Display}} = 1 | Z = 0\right] \cdot \left(1 - \bar{\theta}_{\text{Direct Traffic}}\right)} \tag{17}$$

$$\underline{\delta}_{\text{Email}} = 1 - \frac{E\left[V^{obs}_{\text{Email}} = 1 | Z = 1\right]}{E\left[V^{obs}_{\text{Email}} = 1 | Z = 0\right] \cdot \left(1 - \bar{\theta}_{\text{Direct Traffic}}\right)} \tag{18}$$

Proposition 2 follows by noting that the upper bound for the consent effect on the direct traffic attribution channel occurs when $\underline{\delta}_{\text{Direct Traffic}} = 0$. Then, Assumption 5 allows us to substitute $\bar{\theta}_{\text{Direct Traffic}} = \bar{\theta}_{\text{Display}} = \bar{\theta}_{\text{Email}}$ into the channel analogue of equation (13).

Table 8 presents the results of jointly estimating $\left(\bar{\theta}_{\text{Direct Traffic}}, \underline{\delta}_{\text{Display}}, \underline{\delta}_{\text{Email}}\right)$ using GMM for the display and email marketing attribution channels. Table 8 also includes the corresponding marginal effect estimates for each $\underline{\delta}$ in pageviews and revenue terms, respectively. To construct lower bound marginal effect estimates, we multiply the $\left(\underline{\delta}_{\text{Display}}, \underline{\delta}_{\text{Email}}\right)$ estimates by their corresponding channel share from Table 2. We then multiply this marginal effect by the median firm's weekly pre-GDPR outcomes for additional context.

Table 8: Marketing effect estimates: last-touch attribution sample

Channel	Display	Email	Total
Pageviews			
Channel-level real effect lower bound ($\underline{\delta}_A$)	-26.8%	-4.4%	-
Lower bound marginal effect	-0.3%	-0.2%	-0.5%
Lower bound marginal effect: median firm, weekly	-871.6	-427.8	-1,299.4
Revenue			
Channel-level real effect lower bound ($\underline{\delta}_A$)	-23.3%	-3.5%	-
Lower bound marginal effect	-0.1%	-0.3%	-0.4%
Lower bound marginal effect: median firm, weekly	-$90.71	-$268.80	-$359.51

Note: The marketing marginal effect is computed relative to the weekly outcomes of the respective median last-touch attribution dashboard pre-GDPR.

For both pageviews and revenue, we estimate that the display and email channels are negatively impacted by the GDPR. For pageviews, across both channels, we estimate a real decrease of at least 1,299 pageviews per week for the median firm. This marketing effect lower bound represents a real decline of 0.5%, which is 7.0% of our last-touch attribution aggregate point estimate (see Section 4.3.1). Most of this decline is due to the display ad channel. For the e-commerce sample, we estimate a real loss of at least $360 per week for the median firm. This marketing effect lower bound represents a 0.4% drop in real revenue, or 4.7% of our last-touch attribution sample's aggregate estimate.

Our marketing effect lower-bound estimates are conservative in several respects. We rely on last-touch attribution measures of marketing effectiveness that ignore the cumulative effect of marketing over time. Furthermore, these measures ignore cross-channel spillovers in general and the spillovers from display and email advertising to direct traffic in particular.

6 Conclusion

Online firms use personal data to analyze consumer behavior, market themselves, and generate revenue. Privacy regulation limits personal data use with important consequences for online firms. We study the economic consequences of the GPDR for a large and diverse collection of online firms. Our Adobe Analytics data reveal the website performance of 1,084 firm dashboards. Relative to the previous year, we show that recorded pageviews fall by 11.7% and e-commerce recorded revenue falls 13.3% from EU users after the GDPR. However, these data alone do not distinguish between the real and recording effects on the GDPR. We propose a model to separate the GDPR's real effect on the volume of site visits and the GDPR's consent effect on the recording of site visit outcomes. We conclude that consent accounts for at least 4.7% of the recorded pageview estimate. We also provide conservative estimates for the contribution of GDPR's real effect on personalized marketing. The marketing effect alone represents 7.0% of the recorded pageview estimate and 4.6% of the recorded revenue estimate.

We acknowledge important limitations to our research. First, while our data captures a significant portion of the online economy, the firms that use Adobe Analytics are selected—especially those that also track last-touch attribution data. Second, the GDPR's global repercussions make selecting a control group challenging, though we show our findings are robust to alternative control group choices. Third, we cannot separate the consent and real effects of the GDPR without making assumptions on how they differentially affect recorded site analytics data. Subsequent research in this area employs complementary approaches: e.g., identifying website compliance approaches (Lefrere et al., 2020) and examining user panel data (Zhao et al., 2021). Fourth, we analyze the economic consequences for firms, but do not provide a full welfare analysis of the GDPR. This is a general challenge in privacy settings due to the "privacy paradox" (Barnes, 2006; Athey et al., 2017) whereby individual stated and revealed preferences for privacy diverge. Here, we lack consumer-level decision data to identify the demand for privacy. Fifth, we do not undertake a full profit analysis in this paper as we do not observe costs. Finally, we evaluate the early impact of the GDPR as it was interpreted by online firms in 2018. This period featured both limited regulatory enforcement and website compliance efforts that EU regulators broadly considered to be inadequate (Autoriteit Persoonsgegevens, 2019; Data Protection Commission, 2020b). Our results reflect this reality.

Our study offers several takeaways for privacy regulators. Data minimization is a key principle of the GDPR. EU regulators believe that the collection of web analytics data can pose a privacy risk to EU users and EU regulators therefore want websites to obtain GDPR-compliant consent from users. We document modest progress towards this goal and bound the average non-consent rate to be between [4.0%, 12.8%] for our full sample and [3.6%, 13.7%] for our e-commerce sample. These results are consistent with the prevailing industry practice in 2018 whereby websites applied a *de facto* opt-out approach to consent (Utz et al., 2019). Despite concerns of consent fatigue (European Data Protection Board, 2020), a substantial minority of EU users apparently make the effort to register their non-consent preferences. We provide evidence from our e-commerce dashboards that smaller firms obtain lower consent rates, which suggests the GDPR may have consequences for competition.

Despite evidence of incomplete compliance, our results illuminate real consequences of the GDPR for online firms. First, we show larger drops in pageviews and revenue from users who click on a display ad or a marketing email—channels that rely on personal data. After adjusting for consent, we provide conservative estimates for the GDPR's real effect on outcomes generated via the display and email marketing channels: for the median firm, we estimate a real decrease of at least 0.5% in weekly pageviews in the full sample and 0.4% in weekly revenue in the e-commerce sample. Second, sites that show ads will lose personalized ad revenue from users who do not consent to data processing. Our back-of-the-envelope calculations put these additional costs between 0.3% and 6.1% of ad revenue. Third, limiting analytics data may hamper online firm's ability to create value from this data. This concern is limited by low non-consent rates on average but could become acute under stricter enforcement.

References

Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association 105*(490), 493–505.

Adjerid, I., A. Acquisti, R. Telang, R. Padman, and J. Adler-Milstein (2016). The impact of privacy regulation and technology incentives: The case of health information exchanges. *Management Science 62*(4), 1042–1063.

Adobe (2019). Adobe Analytics Documentation.

Adobe (2021). 2021 digital economy index. https://business.adobe.com/resources/digital-economy-index.html.

Aridor, G., Y.-K. Che, W. Nelson, and T. Salz (2020). The economic consequences of data privacy regulation: Empirical evidence from GDPR. *Available at SSRN*.

Article 29 Data Protection Working Party (2012, June). Opinion 04/2012 on cookie consent exemption. Technical report, Article 29 Data Protection Working Party.

Athey, S., C. Catalini, and C. Tucker (2017). The digital privacy paradox: Small money, small costs, small talk.

Autoriteit Persoonsgegevens (2019, December). Ap: veel websites vragen op onjuiste wijze toestemming voor plaatsen tracking cookies.

Baicker, K. and T. Svoronos (2019, July). Testing the validity of the single interrupted time series design. Working Paper 26080, National Bureau of Economic Research.

Barnes, S. B. (2006). A privacy paradox: Social networking in the united states. *First Monday 11*(9).

Berman, R. and A. Israeli (2022, March). The added value of descriptive analytics: Evidence from online retailers. *Marketing Science*.

Bresnahan, T. F., E. Brynjolfsson, and L. M. Hitt (2002, February). Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence. *The Quarterly Journal of Economics 117*(1), 339–376.

Brynjolfsson, E., L. M. Hitt, and H. H. Kim (2011). Strength in numbers: How does data-driven decision-making affect firm performance? *SSRN eLibrary*.

Brynjolfsson, E. and J. Oh (2012). The attention economy: Measuring the value of free digital services on the internet. In *ICIS*.

Budak, C., S. Goel, J. Rao, and G. Zervas (2016). Understanding emerging threats to online advertising. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pp. 561–578. ACM.

Campbell, J., A. Goldfarb, and C. Tucker (2015). Privacy regulation and market structure. *Journal of Economics & Management Strategy 24*(1), 47–73.

Cavallo, A. and R. Rigobon (2016, May). The billion prices project: Using online prices for measurement and research. *Journal of Economic Perspectives 30*(2), 151–78.

Chen, Z. (2022). Privacy costs and consumer data acquisition: An economic analysis of data privacy regulation. *Available at SSRN 4085923*.

Commission Nationale de l'Informatique et des Libertés (2020a, October). Cookies et autres traceurs : la cnil publie des lignes directrices modificatives et sa recommandation.

Commission Nationale de l'Informatique et des Libertés (2020b, September). Délibération 2020-091 du 17 septembre 2020. Technical report, Commission Nationale de l'Informatique et des Libertés.

Competition and Markets Authority (2020). Appendix F: the role of data in digital advertising. In *Online platforms and digital advertising market study.*

Conley, T. G. and C. R. Taber (2011, 02). Inference with "Difference in Differences" with a Small Number of Policy Changes. *The Review of Economics and Statistics 93*(1), 113–125.

Data Protection Commission (2020a, April). Guidance note: Cookies and other tracking technologies. Technical report, Data Protection Commission.

Data Protection Commission (2020b, April). Report by the data protection commission on the use of cookies and other tracking technologies. Technical report, Data Protection Commission.

Dolfen, P., L. Einav, P. J. Klenow, B. Klopack, J. D. Levin, L. Levin, and W. Best (2019, February). Assessing the gains from e-commerce. Working Paper 25610, National Bureau of Economic Research.

Doudchenko, N. and G. W. Imbens (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Working Paper 22791, National Bureau of Economic Research.

European Commission (2008, January). Flash eurobarometer 226: Data protection in the european union : Data controllers' perceptions. https://data.europa.eu.

European Commission (2019, July 24). Data protection rules as a trust-enabler in the eu and beyond – taking stock. Communication from the commission to the european parliament and the council, European Commission.

European Data Protection Board (2020, May). Guidelines 05/2020 on consent under regulation 2016/679. Technical report, European Data Protection Board.

EuroStat (2020, April). Online sales continue to grow for EU enterprises.

Forrester (2017, December). The forrester wave: Web analytics, q4 2017. Online Report.

Friedman, J. H., T. Hastie, and R. Tibshirani (2010). Glmnet: Lasso and elastic-net regularized generalized linear models. Technical report, http://CRAN.R-project.org/package=glmnet.

Godinho de Matos, M. and I. Adjerid (2021). Consumer consent and firm targeting after GDPR: The case of a large telecom provider. *Management Science*.

Goldfarb, A. and C. Tucker (2011). Privacy regulation and online advertising. *Management Science 57*(1), 57–71.

Goolsbee, A. D. and P. J. Klenow (2018). Internet rising, prices falling: Measuring inflation in a world of e-commerce. In *AEA Papers and Proceedings*, Volume 108, pp. 488–92.

Information Commissioners Office (2019). Cookies – what does 'good' look like?

Information Commissioner's Office (2019, October). Disclosure irq0873632.

Janssen, R., R. Kesler, M. E. Kummer, and J. Waldfogel (2022). GDPR and the lost generation of innovative apps. Technical report, National Bureau of Economic Research.

Jia, J., G. Z. Jin, and L. Wagman (2020). GDPR and the localness of venture investment. *Available at SSRN 3436535*.

Jia, J., G. Z. Jin, and L. Wagman (2021). The short-run effects of the General Data Protection Regulation on technology venture investment. *Marketing Science 40*(4), 661–684.

Johnson, G., S. Shriver, and S. Du (2020). Consumer privacy choice in online advertising: Who opts out and at what cost to industry? *Marketing Science 39*(1), 33–51.

Johnson, G. A. (2022). Economic research on privacy regulation: Lessons from the GDPR and beyond. Work in progress.

Johnson, G. A., S. K. Shriver, and S. G. Goldberg (2022). Privacy & market concentration: intended & unintended consequences of the GDPR. Available at SSRN.

Jones, M. L. and M. E. Kaminski (2020). An american's guide to the GDPR. *Denver Law Review 98*(1).

Ke, T. T. and K. Sudhir (2020). Privacy rights and data security: GDPR and personal data driven markets. *Available at SSRN 3643979*.

Koski, H. and N. Valmari (2020). Short-term impacts of the GDPR on firm performance. ETLA Working Papers.

Lambrecht, A., K. Seim, and C. Tucker (2011). Stuck in the adoption funnel: The effect of interruptions in the adoption process on usage. *Marketing Science 30*(2), 355–367.

Lefrere, V., L. Warberg, C. Cheyre, V. Marotta, and A. Acquisti (2020). The impact of the GDPR on content providers.

Libert, T., L. Graves, and R. K. Nielsen (2018). Changes in third-party content on European news websites after GDPR.

Long, L. (2020, January). How to a/b test to optimize the data collection consent experience for users.

Manski, C. F. (2005). Partial identification with missing data: concepts and findings. *International Journal of Approximate Reasoning 39*(2), 151 – 165. Imprecise Probabilities and Their Applications.

McDonald, A. M., R. W. Reeder, P. G. Kelley, and L. F. Cranor (2009). A comparative study of online privacy policies and formats. In I. Goldberg and M. J. Atallah (Eds.), *Privacy Enhancing Technologies*, Berlin, Heidelberg, pp. 37–55. Springer Berlin Heidelberg.

Miller, A. R. and C. Tucker (2009). Privacy protection and technology diffusion: The case of electronic medical records. *Management Science 55*(7), 1077–1093.

Paz, L. S. and J. E. West (2019, June). Should we trust clustered standard errors? a comparison with randomization-based methods. Working Paper 25926, National Bureau of Economic Research.

Peukert, C., S. Bechtold, M. Batikas, and T. Kretschmer (2022). Regulatory spillovers and data governance: Evidence from the gdpr. *Marketing Science forthcoming*.

Prasad, A. and D. R. Perez (2020). The effects of GDPR on the digital economy: Evidence from the literature. *Informatization Policy 27*(3), 3–18.

PricewaterhouseCoopers (2018). Pulse survey: GDPR budgets top $10 million for 40% of surveyed companies. https://www.pwc.com/us/en/services/consulting/library/general-data-protection-regulation-gdpr-budgets.html.

Quantcast (2018, July). Quantcast choice powers one billion consumer consent choices in two months since gdpr. Press Release.

Ravichandran, D. and N. Korula (2019, August). Effect of disabling third-party cookies on publisher revenue. Technical report, Google Inc.

Sanchez-Rola, I., M. Dell'Amico, P. Kotzias, D. Balzarotti, L. Bilge, P.-A. Verviker, and I. Santos (2019). Can I opt out yet? GDPR and the global illusion of cookie control. In *ACM ASIACCS 2019*.

Schmitt, J., K. M. Miller, and B. Skiera (2021). The impact of privacy laws on online user behavior. *arXiv preprint arXiv:2101.11366*.

Snelders, E., L. Worp, and S. Song (2020). A future without advertising cookies? It's possible! Technical report, Ster.

Sørensen, J. and S. Kosta (2019). Before and after GDPR: The changes in third party presence at public and private european websites. In *The World Wide Web Conference*, WWW '19, New York, NY, USA, pp. 1590–1600. ACM.

TrustArc (2018, July). GDPR compliance status: A comparison of us, uk and eu companies. Technical report, TrustArc.

U.S. Census Bureau (2019). 2017 e-stats report: Measuring the electronic economy.

Utz, C., M. Degeling, S. Fahl, F. Schaub, and T. Holz (2019). (un) informed consent: Studying GDPR consent notices in the field. In *Proceedings of the 2019 acm sigsac conference on computer and communications security*, pp. 973–990.

Zhao, Y., P. Yildirim, and P. K. Chintagunta (2021). Privacy regulations and online search friction: Evidence from GDPR. *Available at SSRN 3903599*.

Zhuo, R., B. Huffaker, kc claffy, and S. Greenstein (2021). The impact of the General Data Protection Regulation on internet interconnection. *Telecommunications Policy 45*(2), 102083.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology) 67*(2), 301–320.

Table 9: GDPR main effect estimates: panel difference estimator

	(2)
Sample	E-commerce
Dependent variable†	log(Pageviews + 1)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post-GDPR}\}$	-0.118
	(0.030)
$\mathbb{1}\{2018\}$	0.111
	(0.022)
Average marginal effect	-11.15
RSID + Week FE	Y
R^2	0.969
N	22,436

Note: Standard errors clustered at the Dashboard + Week level. †Recorded outcomes.

Online Appendices

A Main effect robustness

We first provide pageviews results for the e-commerce sample in Section A.1. To address potential threats to validity, we outline alternative empirical strategies. Section A.2.1 addresses a concern about anticipatory or delayed compliance behavior. Section A.2.2 outlines a synthetic controls approach that relaxes an underlying restriction in our panel difference estimator and allows us to more flexibly construct a control group. Section A.2.3 discusses an alternate control group and two associated model specifications that account for any contemporaneous shocks to the web outcomes of Western users. Section A.3 provides the results for these robustness exercises.

A.1 Pageviews results for the e-commerce sample

Table 9 provides the pageviews outcome estimates for the e-commerce site sample. The panel differences estimation approach mirrors that in our main effect estimates Table 3 and the resulting estimates are comparable to both the full sample pageviews estimate and the e-commerce sample revenue estimate.

A.2 Alternative empirical strategies

A.2.1 Window regressions (PD-WR)

To address concerns about anticipatory or delayed compliance by websites, we remove a two-month window around the GDPR from our data and re-estimate the regression in equation (3). Firms made large investments to comply with the GDPR. Surveys reveal that some firms completed this work before the enforcement deadline while other firm's efforts were ongoing (TrustArc, 2018). Since firms can make quick changes to their website and online marketing, they have an incentive to wait until the last minute to implement their online GDPR compliance changes. Other research confirms that most websites waited until the enforcement deadline before making changes to their sites (e.g. Sørensen and Kosta 2019; Johnson et al. 2022). Note also that Figure 2 displays no change in trend before the GDPR took effect. This is further supported by pre-enforcement placebo checks in Online Appendix A.4. Nevertheless, anticipatory or delayed compliance relative to the deadline may lead us to under- or over-estimate the effect of the GDPR in equation (3). We address this by reestimating equation (3) after dropping four weeks of data both before and after the deadline: we term these the "window regression" results.

A.2.2 Synthetic controls (SC)

Our empirical strategy's within-dashboard design uses the 2017 dashboards as the control group to capture firm-specific seasonality and characteristics. However, the difference-in-difference model assumes that the counterfactual trend is best represented by an equal weighting of these dashboards in 2017. Synthetic controls present a data-driven alternative for selecting the control group. We use the synthetic controls approach (Abadie et al., 2010; Doudchenko and Imbens, 2016) to construct a re-weighted control group that relaxes our within-dashboard restriction in order to predict the counterfactual in the post-GDPR period.

Synthetic controls flexibly construct a control group by taking a weighted average of control-units in order to best predict outcomes for treated units. Intuitively, if we can construct a control group that behaves similarly enough to the treatment group in the pre-period, then this control group should behave similarly to how the treatment group would have behaved after the intervention date, had it not received treatment. We define a *control-unit* to be any 2017 dashboard (logged) data and our *treated-unit* to be the mean of our (logged) 2018 dashboard data, plotted in Figure 7. Because we have one control-unit per dashboard (1,084 in our *full* sample and 353 in our *e-commerce* sample) and only one treated unit, we follow Doudchenko and Imbens (2016) and use an elastic net to construct our synthetic control. Weights are chosen in order to minimize the pre-treatment difference between the treated-unit and potential control units. The intent of synthetic controls is to predict the counterfactual, thus we use cross-validation to incorporate prediction error into our objective function. Then, the counterfactual is constructed by taking the chosen weights and using them to aggregate post-treatment control-unit outcomes. We can then recover the treatment effect by differencing the treated and synthetic control outcomes. Online Appendix A.5.2 details the cross-validation and model fitting procedure.

A.2.3 North American difference-in-differences (DD-NA) and triple panel difference (TPD)

We next consider a contemporaneous control group. Our 2017 control group would not account for any 2018 contemporary confounds like global macroeconomic changes. We thus consider the 2018 web outcomes of North American users from our dashboard sample as an alternate control group. Table 1 reveals that our dashboards have slightly less North American traffic than EU traffic, on average. E-commerce dashboards accrue roughly equal percentages of their revenue from the EU and North America. We explain above that this control group is likely contaminated—owing to within-firm spillovers of GDPR compliance to non-EU users. Thus, any specification with North America as the control group may understate the effect of the GDPR. Other GDPR studies use similar contemporaneous controls (Jia et al., 2021; Aridor et al., 2020; Zhuo et al., 2021), though these authors also acknowledge this contamination issue. We present both our panel estimator with North American controls and a triple panel difference specifications using the North America control. Both specifications address a confounding and contemporaneous shock to online outcomes to both EU and North American users. The triple panel difference specification compares our preferred EU panel difference estimate (equation 3) and an analogous North American panel difference estimate that uses outcomes from 2017 and 2018. The triple panel difference specification identifies the GDPR effect separately from continent-specific seasonality (shared in 2017 and 2018) and a common shock after May 2018 to both EU and North American users.

A.3 Robustness results

In this section, we present the results of our robustness tests. First, in Table 10 we present our panel differences model with alternative fixed effects specifications including (1) a saturated model with dashboard-by-week fixed effects (columns (2) and (4)) and (2) a model that allows for dashboard specific linear time trends. In addition, Table 10 includes the results of our panel differences window regression specification from Section A.2.1. Our main effect

Table 10: Robustness: alternative fixed effects specifications and window regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	PD - Preferred	PD	PD	PD-WR	PD - Preferred	PD	PD	PD-WR
Sample	All Dashboards				E-commerce			
Dependent variable†	log(Pageviews + 1)				log(Revenue + 1)			
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post-GDPR}\}$	-0.124	-0.124	-0.137	-0.174	-0.142	-0.138	-0.095	-0.176
	(0.026)	(0.026)	(0.022)	(0.025)	(0.030)	(0.03)	(0.028)	(0.031)
$\mathbb{1}\{2018\}$	0.050	0.050		0.056	0.193	0.192		0.215
	(0.014)	(0.014)		(0.015)	(0.028)	(0.027)		(0.028)
Dashboard + Week FE	Y	N	N	Y	Y	N	N	Y
Dashboard x Week FE	N	Y	N	N	N	Y	N	N
Dashboard x Trend	N	N	Y	N	N	N	Y	N
R^2	0.969	0.983	0.738	0.966	0.962	0.983	0.734	0.961
N	69,344	69,344	69,344	49,832	22,436	22,436	22,436	16,111

Note: Standard errors clustered at the Dashboard + Week level. †Columns (1) and (5) are the preferred specifications from Table 3. Columns (2) and (6) present alternative fixed effects specifications. Columns (3) and (7) instead employ dashboard-specific linear time trends. Columns (4) and (8) implement the window regression specification detailed in Section A.2.1.

results are largely unchanged by the addition of richer fixed effects. Allowing for dashboard-specific, linear-growth trends (columns (3) and (7)) leads to a slight increase (decrease) in our estimated treatment effect for the pageviews (revenue) outcome, though the interpretation and significance of our results remains unchanged. Finally, both our window regression estimates (columns (4) and (8)) are higher than our preferred estimates in Table 3, though they remain within the original confidence intervals.[20] The marginal effects here are -16.0% for recorded pageviews and -16.1% for recorded revenue. These higher estimates could arise if several websites delayed their compliance with the GDPR.

Table 11 presents three alternative research designs as described above: difference-in-differences with North America as the control group (denoted by DD-NA), triple differences (TPD), a triple differences specification with rich fixed effects (TPD-FE), and synthetic controls (SC).

Table 11 indicates that our main effect results are robust to several alternate specifications. Both specifications using North America as a control group yield smaller point

[20] To address the possibility of further delayed compliance, we also considered models that instead dropped 5, 6, 7, and 8 weeks of data after the GDPR (as well as 4 weeks before). These alternative specifications had little impact on the window regression results. We thank a reviewer for suggesting this analysis. Results available upon request.

estimates though each remains negative and statistically significant at the 1% level. Of the two specifications, the more conservative marginal effect estimates are -5.4% for recorded pageviews and -5.6% for recorded revenue. We take these lower point estimates as evidence of spillover effects from the GDPR on North American traffic, as discussed in Section 4. Columns (4) and (8) build off the triple panel differences design in columns (3) and (5) by including week-by-year fixed effects. Finally, our synthetic control results in columns (5) and (10) indicate marginal effects of -8.7% for recorded pageviews though only -1.4% for recorded revenue. Inference using synthetic controls is difficult: we follow placebo procedures outlined in Abadie et al. (2010). Only 0.2% of placebos achieve a magnitude of prediction error similar to our pageview result. The revenue synthetic controls estimate is less robust, with 22.4% of placebos achieving a similar magnitude of prediction error. In Online Appendix A.5, we discuss the synthetic control procedure and results in detail.

Taken as a whole, these alternate specifications seek to address potential threats to validity in our preferred empirical approach. The results indicate a robust negative impact of the GDPR on recorded site outcomes, though somewhat less so for our revenue results owing to the synthetic controls analysis.

A.4 Placebo tests

We run placebo tests to assess the potential for false positive effect estimates. We implement placebo tests by first choosing a counterfactual treatment week from the pre-GDPR period of our data. Then, equation (19) is estimated using data from before April 25th (pre-GDPR). We exclude one-month before the implementation of the GDPR in order to omit any anticipatory behavior. This procedure is repeated for placebo treatment dates ranging from 14 to 7 weeks prior to May 25th, for a total of 8 placebo tests. These placebo dates are chosen in order to provide adequate pre-trends and post-trends in the data (at least 3 data points before and after the placebo treatment).

$$\log\left(y_{itw}+1\right) = \alpha\mathbb{1}\{2018\}_t + \beta_p\left(\mathbb{1}\{2018\}_t \text{ x } \mathbb{1}\{\text{Post Placebo}\}_w\right) + \theta_i + \eta_w + \epsilon_{itw} \tag{19}$$

The primary coefficient of interest is β_p. Fixed effects are included as in equation (3) and all standard errors are clustered at the dashboard-week level. Significant point estimates are indicative of false positives; a prevalence of false positives may undermine the credibility of our point estimates.

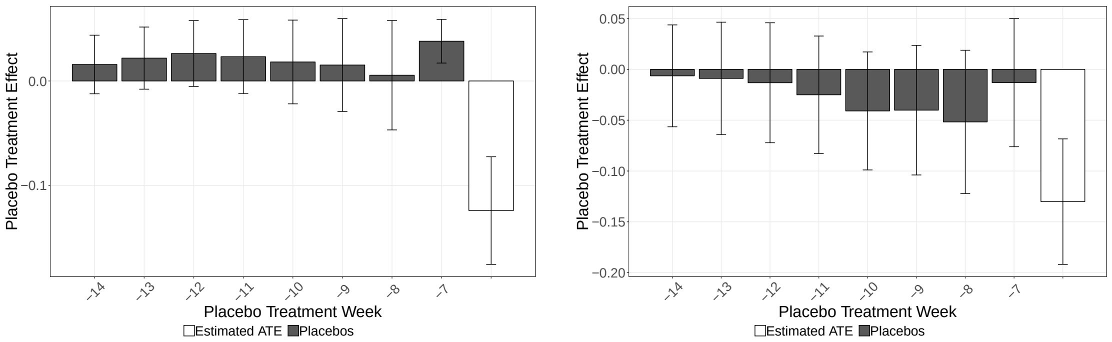
Our placebo results are presented in Figure 6 in grey. Figure 6 also includes our estimated treatment effect in white, to aid in comparison. The placebo results demonstrate the robustness of our identification strategy and point estimates in Table 3. The placebo

Table 11: Robustness: alternative research designs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Model†	PD - Preferred	DD-NA	TPD	TPD-FE	SC	PD - Preferred	DD-NA	TPD	TPD-FE	SC
Dependent variable††	log(Pageviews + 1)					log(Revenue + 1)				
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post-GDPR}\}$	-0.124		-0.067			-0.142		0.028		
	(0.026)		(0.024)			(0.030)		(0.040)		
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post-GDPR}\}\times\mathbb{1}\{\text{EU}\}$			-0.056	-0.055				-0.149	-0.150	
			(0.019)	(0.019)				(0.049)	(0.049)	
$\mathbb{1}\{\text{EU}\} \times \mathbb{1}\{\text{Post-GDPR}\}$		-0.084	-0.036	-0.036			-0.058	0.097	0.097	
		(0.005)	(0.002)	(0.001)			(0.013)	(0.010)	(0.008)	
$\mathbb{1}\{2018\}$	0.050		0.076			0.193		0.050		
	(0.014)		(0.017)			(0.028)		(0.037)		
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{EU}\}$			-0.026	-0.026				0.144	0.144	
			(0.016)	(0.016)				(0.041)	(0.041)	
Average marginal effect	-11.67%	-8.09%	-5.44%	-5.44%	-8.70%	-13.26%	-5.59%	-13.87%	-13.89%	-1.40%
Dashboard + Week FE	Y	N	N	N		Y	N	N	N	
Dashboard + Week + Region FE	N	Y	Y	N		N	Y	Y	N	
Dashboard + Week x Region FE	N	N	N	Y		N	N	N	Y	
R^2	0.969	0.484	0.491	0.491		0.962	0.391	0.397	0.397	
N	69,344	64,931	138,280	138,280		22,436	20,199	42,721	42,721	

Note: Standard errors clustered at the dashboard + week level. †Model: 1) DD-NA: North American user outcome control; 2) TPD: triple panel differences; 3) TPD-FE: triple panel differences with week-by-region fixed effects; and 4) SC: synthetic controls. ††Recorded outcomes.

Figure 6: Main effect panel difference pre-trend placebo tests


(a) Log recorded pageviews

(b) Log recorded revenue

Note: Estimates of placebo treatment effects, β_p from equation (19), with 95% confidence intervals.

estimates are smaller in magnitude than our main results and have the opposite sign. All but one estimate—seven weeks prior for the pageviews outcome—are statistically insignificant.

A.5 Synthetic controls: additional details

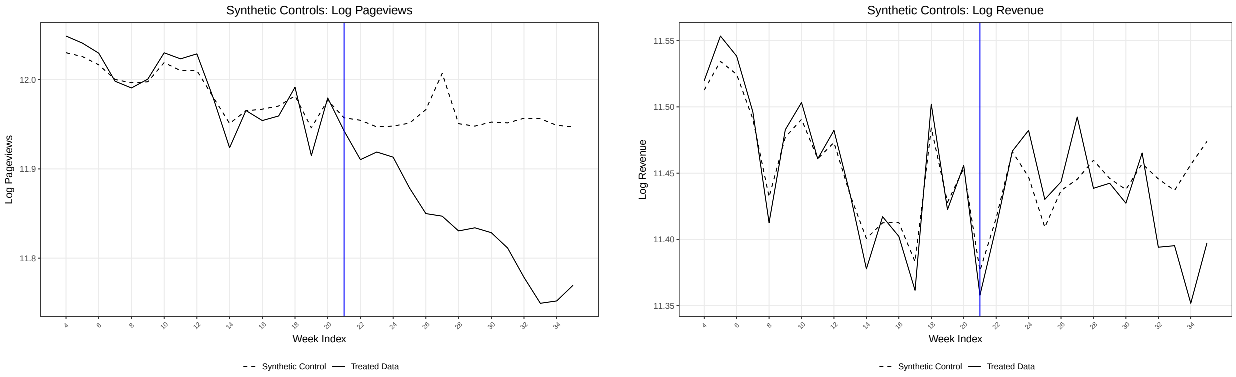
A.5.1 Synthetic controls: detailed results

Section A.2.2 discusses the motivation for implementing synthetic controls. Here we discuss the results and methodology behind this approach in more detail.

Figure 7 plots average outcomes for our treated group with solid line and our fitted control group in a dotted line. The vertical line marks the implementation of the GDPR. The details of constructing our control group can be found in Section A.5.2. For both pageviews and revenue, Figure 7 illustrates a well fitted control group in the pre-GDPR period. For pageviews, the predicted control group deviates from our 2018 outcomes in the post-GDPR period - congruent with Figure 2. In contrast, our predicted control group continues to match the treated group for revenue well into the post-GDPR period. Our point estimates for each outcome are presented in Table 11. The synthetic controls method estimates a treatment effect of 8.7% for pageviews and 1.4% for revenue.

Constructing standard errors in a synthetic controls setting with an elastic net is not a well understood problem. To provide some sense of the robustness of our synthetic controls results, we appeal to Abadie et al. (2010). Abadie et al. (2010) implement placebo studies by asking "how often would we obtain results of this magnitude if we had chosen a random (counterfactual) treated unit, rather than the factual treated unit?" In our context, we select a placebo treated unit from our control unit, fit a control group using our synthetic

Figure 7: Synthetic controls results: fitted trends

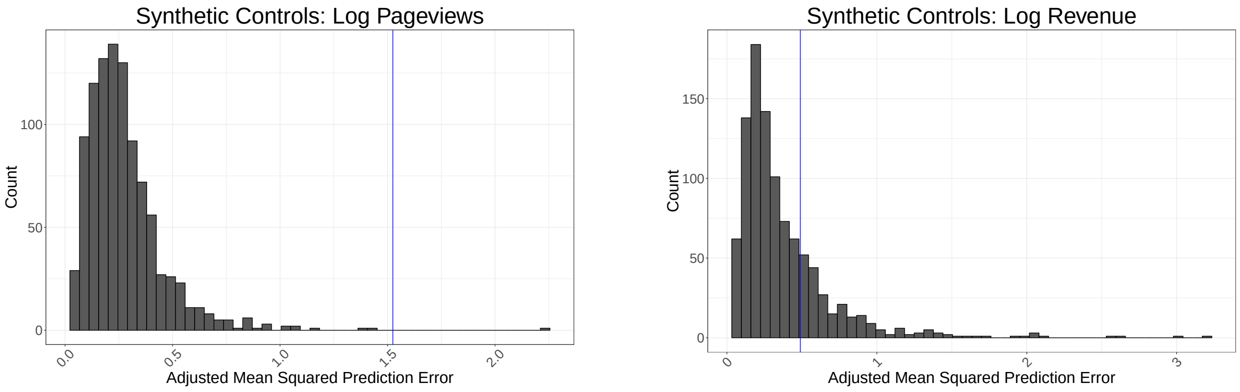


(a) Log recorded pageviews

(b) Log recorded revenue

Note: This figure presents the fitted synthetic control groups constructed using the procedure outlined in Online Appendix A.5.

Figure 8: Synthetic control results: placebo tests



(a) Log recorded pageviews

(b) Log recorded revenue

Note: This figure presents the distribution of AMPSE estimates from 1,000 synthetic control placebo studies. Blue lines indicate the true estimated AMPSE.

controls procedure, and then estimate a placebo treatment effect. We detail this procedure in Section A.5.3. In comparing the placebo and factual synthetic controls, we follow Abadie et al. (2010) in using an adjusted mean squared prediction error statistic (AMSPE).

Figure 8 presents histograms of our 1000 placebo studies for pageviews and revenue. The histograms reflect the distribution of AMSPE estimates across our 1000 placebo trials, for each outcome. The vertical line marks the AMSPE of our true estimated treatment effect, 1.5 for pageviews and 0.49 for revenue. For pageviews, our true AMPSE is at the 99.8th quantile of the placebo AMPSE distribution, which suggests that recovering a treatment effect of the magnitude presented in Table 11 column (4) by chance is highly unlikely. The AMPSE for revenue is at the 78.4th quantile of the placebo AMPSE distribution. Our synthetic controls estimate for revenue is therefore less precise than our pageviews measure, and somewhat more likely to occur by chance.

A.5.2 Cross-validation & estimation

For the following discussion, T_0 will be the period before which the intervention takes place, $Y(0)$ is the counterfactual outcome, and $Y(1)$ is the observed outcome of the treatment. In our setting, we will use 2017 dashboard log outcome data as control units and the average (across dashboards) of 2018 log outcome data as our treatment unit. That is, for each outcome variable, we have a treatment unit and a set of control units, denoted by C:[21]

$$Y_t = \frac{1}{N}\sum_{i}^{N} \log\left(y_{it}^{2018} + 1\right) \tag{20}$$

$$C = \{C_{it} = \log\left(y_{it}^{2017} + 1\right) \forall\, i\} \tag{21}$$

We estimate weights such that the weighted combination of C_{it} best matches Y_t. In our setting, we have 17 pre-treatment time periods and 1084 control units, or $N >> T_0$. We follow Doudchenko and Imbens (2016) in using an elastic net to construct our control group. See Zou and Hastie (2005) for a detailed discussion of elastic nets and their properties. In brief, we fit a model with the following objective function:

$$Q(\mu, \omega | Y_t, C_{it}; \alpha, \lambda \text{ for } t < T_0) = ||Y_t - \mu - \omega C_{it}||_2^2 + \lambda \cdot \left(\frac{1-\alpha}{2}||\omega||_2^2 + \alpha||\omega||_1\right) \tag{22}$$

Where μ is a constant, ω is a vector of length N of weights, and α and λ are penalty parameters chosen by the econometrician.

We choose penalty parameters using a modified version of the cross validation routine proposed in Doudchenko and Imbens (2016). In particular, for a proposed pair of penalty parameters, $\{\alpha', \lambda'\}$, we construct pseudo treated units as follows. First, we partition C into B random partitions of size b. We will refer to a partition as C^b. Each C^b is used to construct a pseudo treated unit, $Y_t^{C^b}$, by taking the average over units $i \in C^b$. We use $\tilde{C} = C \backslash C^b$ as the control units for pseudo treated unit $Y_t^{C^b}$. An elastic net is fitted, using only *pre-intervention* data, to obtain $\{\hat{\mu}^b, \hat{\omega}^b\}$. That is: $\{\hat{\mu}^b, \hat{\omega}^b\} = \text{argmin}_{\mu,\omega} \sum_{t=1}^{T_0} \left(Y_t^{C^b} - \mu - \omega\tilde{C}_{it}\right)^2 + \lambda' \cdot \left(\frac{1-\alpha'}{2}||\omega||_2^2 + \alpha'||\omega||_1\right)$. Given the weights estimated above and using the proposed penalty parameters $\{\alpha', \lambda'\}$, we predict the outcome for $Y_t^C(0)$ in $t > T_0$ and construct the mean

[21] Note that we are reusing notation in this section. For instance, C refers to the control group here and not the consent potential outcome as elsewhere.

squared error for each B.

$$Y_t^{C^b}(0) = \hat{\mu}^b + \hat{\omega}^b \tilde{C}_{it} \tag{23}$$

$$CV_B(\alpha', \lambda') = \frac{1}{T - T_0} \sum_{t=T_0}^{T} \left(Y_t^{C^b}(1) - Y_t^{C^b}(0) \right)^2 \tag{24}$$

Model performance is then evaluated using the average of the cross validated mean squared error across our B partitions:

$$CV(\alpha', \lambda') = \frac{1}{B} \sum_b CV_b(\alpha', \lambda') \tag{25}$$

Finally, tuning parameters are chosen such that $\{\alpha, \lambda\} = argmin_{\alpha', \lambda'} CV(\alpha', \lambda')$. Using these tuning parameters, we recover the vector of weights ω needed to construct our synthetic control.

We search over a grid of $\alpha \in [.01, .99]$ in increments of .01 and take advantage of the λ validation built into the *glmnet* package in *R* (Friedman et al., 2010). For each $\{\alpha', \lambda'\}$ we partition the control units into $B = 10$ samples—analogous to 10 cross-fold validation. We repeat the above procedure 100 times for each outcome variables to construct a set of ω vectors, which we average to construct our final weights. We then generate our estimates of the treatment effect by differencing the average levels of the treated unit and the synthetic control in the post-GDPR period.

A.5.3 Synthetic controls: Placebo routine

We can appeal to Abadie et al. (2010) to get a sense of how reasonable our results are. In particular, the following exercise asks: "how large would our prediction error be had treatment not occurred?" To construct this counterfactual, our procedure is as follows:

- Randomly sample $n = 10$ units from C
- Construct our pseudo-treated unit as $C_t^{pseudo} = \frac{1}{n} \sum_{i \in sample} C_{it}$
- Fit an elastic net to C_t^{pseudo} as described in Online Appendix A.5.2
- Calculate the adjusted mean squared prediction error (Abadie et al., 2010):

$$\frac{T_0}{T - T_0} \frac{\sum_{t=T_0}^{T} \left(Y_t^{pseudo}(1) - Y_t^{C(pseudo)}(0) \right)^2}{\sum_{t=0}^{T_0} \left(Y_t^{pseudo}(1) - Y_t^{C(pseudo)}(0) \right)^2} \tag{26}$$

That is, we calculate the mean squared prediction error and scale it by the mean squared fitting error

- Repeat 1000 times for each outcome

The results of this procedure are presented in Figure 8.

B GDPR effect heterogeneity

B.1 Heterogeneity results

B.1.1 Last-touch attribution regression estimates

We implement the following regression equation to estimate marketing channel-specific GDPR effects:

$$\begin{aligned}\log\left(y_{itwc}+1\right) = &\sum \alpha_c \left(\mathbb{1}\{\text{Post GDPR}\}_w \text{ x } \mathbb{1}\{\text{Channel}\}_c\right) + \\ &\sum \gamma_c \left(\mathbb{1}\{2018\} \text{ x } \mathbb{1}\{\text{Channel}\}_c\right) + \\ &\sum \beta_c \left(\mathbb{1}\{2018\} \text{ x } \mathbb{1}\{\text{Post GDPR}\}_w \text{ x } \mathbb{1}\{\text{Channel}\}_c\right) + \eta_{ic} + \xi_w + \epsilon_{itwc}\end{aligned} \tag{27}$$

where c refers to attribution channel and y_{itwc} denotes recorded outcomes associated with channel c. η_{ic} and ξ_ω are dashboard-channel and week-specific fixed effects, respectively. We interact an indicator for channel c, $\mathbb{1}\{\text{Channel}\}_c$, with each term of equation (3). β_c are the coefficients of interest because they capture the effect of the GDPR by channel c. Under the specified model, the variation identifying each β_c will be differences in outcomes across channels and years in post-May 25 weeks, after accounting for unobservables common to dashboard-channel and week observations, and a common level shift in channel outcomes across years. Table 12 presents the coefficient estimates for the different last-touch attribution channels. Note that these estimates are graphed in Figure 3.

B.1.2 Size heterogeneity results

We examine heterogeneous effects of the GDPR across dashboard size by interacting our main specification (equation 3) with an indicator for large dashboards. This indicator equals one for dashboards with above-median pageviews in the pre-GDPR period. Specifically, we estimate the following regression:

Table 12: Last-touch attribution channel ATE regression estimates

	(1)	(2)
Dependent Variable[†]	log(Pageviews + 1)	log(Revenue + 1)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \mathbb{1}\{\text{Direct Traffic}\}$	-0.089	-0.114
	(0.029)	(0.037)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \mathbb{1}\{\text{Search}\}$	-0.077	-0.052
	(0.027)	(0.035)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \mathbb{1}\{\text{Display}\}$	-0.307	-0.238
	(0.077)	(0.110)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \mathbb{1}\{\text{Email}\}$	-0.109	-0.132
	(0.043)	(0.063)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \mathbb{1}\{\text{Social}\}$	-0.244	-0.042
	(0.039)	(0.063)
RSID + Week FE	Y	Y
R^2	0.950	0.937
N	146,095	52,855

Note: Standard errors clustered at the Dashboard + Week level. [†]Recorded outcomes.

$$\begin{aligned}\log\left(y_{itw}+1\right) = \quad & \alpha_1 \mathbb{1}\{2018\}_t + \beta_1 \left(\mathbb{1}\{2018\}_t \text{ x } \mathbb{1}\{\text{Post-GDPR}\}_w\right) + \\ & \alpha_2 \mathbb{1}\{2018\}_t \text{x} \mathbb{1}\{\text{Large Site}\}_i + \gamma_1 \mathbb{1}\{\text{Post-GDPR}\}_w \text{x} \mathbb{1}\{\text{Large Site}\}_i + \\ & \beta_2 \left(\mathbb{1}\{2018\}_t \text{ x } \mathbb{1}\{\text{Post-GDPR}\}_w \times \mathbb{1}\{\text{Large Site}\}_i\right) + \eta_i + \xi_w + \epsilon_{itw}\end{aligned} \tag{28}$$

The coefficients of interest are β_1, which estimates the average treatment effect of the GDPR on small sites, and the sum of $\beta_1 + \beta_2$ which is an estimate of the average treatment effect of the GDPR on large sites.Table 13 presents the full regression results examining heterogeneity in site usage.

B.1.3 Regulatory Enforcement Regressions

To construct our index of regulatory enforcement beliefs, we use a European Commissions survey from 2008 of 4,835 data controllers across all countries in the EU European Commission (2008). The survey asks to what extent data controllers agree or disagree with “the data protection law in (OUR COUNTRY) is interpreted and applied more rigorously than in other Member States.” Responses are recorded on a four point scale and include a no-response option. We exclude the non-responses and construct a response-weighted average index for each country in the EU that takes values from 0 (all responses are “totally disagree”) to 1 (all responses are “totally agree”). We then standardize the index to have mean of zero and variance of one. We additionally utilize data on income (GDP per capita) for

Table 13: Treatment effect heterogeneity by site size

	(1)	(2)	(3)	(4)
Sample	All Dashboard		E-commerce	
Dependent variable†	log(Pageviews + 1)	Pageviews per Visitor	log(Revenue + 1)	Revenue per Visitor
𝟙{2018} × 𝟙{Post GDPR} × 𝟙{Large Dashboard}	0.040	-0.050	0.101	-0.329
	(0.034)	(0.130)	(0.053)	(0.187)
𝟙{2018} × 𝟙{Post GDPR}	-0.144	0.225	-0.182	0.338
	(0.034)	(0.112)	(0.046)	(0.166)
𝟙{2018} × 𝟙{Large Dashboard}	-0.062	0.176	-0.049	0.399
	(0.024)	(0.124)	(0.046)	(0.188)
𝟙{Post GDPR}× 𝟙{Large Dashboard}	-0.039	0.132	-0.023	0.426
	(0.020)	(0.112)	(0.050)	(0.179)
𝟙{2018}	0.081	-0.379	0.213	-0.167
	(0.018)	(0.113)	(0.043)	(0.177)
Average Marginal Effect				
Small Dashboard	-13.43%	25.77%	-16.69%	40.21%
Large Dashboard	-9.87%	18.79%	-7.85%	0.90%
RSID + Week FE + Size FE	Y	Y	Y	Y
R^2	0.969	0.725	0.960	0.758
N	69,344	69,344	22,812	22,812

Note: Standard errors clustered at the Dashboard + Week level. †Recorded outcomes.

Table 14: Regulatory strictness heterogeneity regressions

	(1)	(2)	(3)	(4)
Dependent Variable†	log(Pageviews + 1)		log(Revenue + 1)	
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \text{Strictness}$	-0.056	-0.040	-0.047	-0.041
	(0.006)	(0.006)	(0.016)	(0.019)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\}$	-0.074	0.041	-0.096	-0.040
	(0.022)	(0.029)	(0.022)	(0.075)
$\mathbb{1}\{2018\} \times \text{Strictness}$	-0.007	-0.008	0.030	0.008
	(0.005)	(0.004)	(0.007)	(0.009)
$\mathbb{1}\{\text{Post GDPR}\}\times\text{Strictness}$	0.032	0.027	0.002	-0.014
	(0.007)	(0.009)	(0.010)	(0.012)
$\mathbb{1}\{2018\}$	0.056	0.049	0.163	-0.025
	(0.008)	(0.024)	(0.015)	(0.059)
$\mathbb{1}\{2018\} \times \text{Income}$		0.019		0.469
		(0.056)		(0.151)
$\mathbb{1}\{\text{Post GDPR}\}\times\text{Income}$		0.101		0.345
		(0.054)		(0.128)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \text{Income}$		-0.294		-0.138
		(0.079)		(0.186)
RSID + Week FE	Y	Y	Y	Y
R^2	0.969	0.969	0.962	0.962
N	69,344	69,344	22,812	22,812

Note: Standard errors clustered at the Dashboard + Week level. †Recorded outcomes.

each EU country in 2018 from the World Bank, which we expect correlates with advertising and e-commerce revenue.[22]

We explore the role of regulatory strictness empirically, by interacting our panel differences estimator with our regulatory strictness measure. We estimate the following equation:

$$\begin{aligned}\log\left(y_{itw}+1\right) = {} & \alpha_1\mathbb{1}\{2018\} + \alpha_2\mathbb{1}\{\text{Strictness x } \mathbb{1}\{2018\} + \alpha_3\mathbb{1}\{\text{Income}\} \text{ x } \mathbb{1}\{2018\}+ \\ & \gamma_1\mathbb{1}\{\text{Strictness}\} \text{ x } \mathbb{1}\{\text{Post GDPR}\}_w + \gamma_2\mathbb{1}\{\text{Income}\}\text{x}\mathbb{1}\{\text{Post GDPR}\}_w + \\ & \beta\mathbb{1}\{\text{Post GDPR}\}_w \text{ x } \mathbb{1}\{2018\}+ \\ & \beta_{income}\mathbb{1}\{\text{Income}\} \text{ x } \mathbb{1}\{\text{Post GDPR}\}_w \text{ x } \mathbb{1}\{2018\}+ \\ & \beta_{stritcness}\mathbb{1}\{\text{Strictness}\} \text{ x } \mathbb{1}\{\text{Post GDPR}\}_w \text{ x } \mathbb{1}\{2018\} + \eta_i + \xi_w + \epsilon_{itw}\end{aligned} \tag{29}$$

Our interaction coefficient of interest is $\beta_{stritctness}$. Following Jia et al. (2021), we also include an interaction with country-level GDP per capita, because income is a potential confound that correlates with regulatory strictness. Results of this regression are presented below.

[22] https://Data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=PL-GR-PT-DE-EU

C Marginal effects

The models in Sections 4 are non-linear and therefore rely on marginal effects for interpretation. In this section we detail the construction of these marginal effects. Our models are generally of the form:

$$log(y_{itw}+1) = \alpha \mathbb{1}\{2018\}_t + \beta \left(\mathbb{1}\{2018\}_t \text{ x } \mathbb{1}\{\text{Post GDPR}\}_w\right) + \eta_i + \xi_w + \epsilon_{itw} \tag{30}$$

where β captures the GDPR effect. First, we estimate the above regression using the data. Then, for the post-GDPR period, we use our estimates to construct predicted outcomes for both the treated and a counterfactual untreated group:

$$y_{itw}^{untreated} = \exp\left(\alpha + \eta_i + \xi_w + \frac{\sigma^2}{2}\right) - 1 \tag{31}$$

$$y_{itw}^{treated} = \exp\left(\alpha + \beta + \eta_i + \xi_w + \frac{\sigma^2}{2}\right) - 1 \tag{32}$$

Where we have included variances ($Var[\epsilon_{itw}] = \sigma^2$) to account for the expected value of the (log-normal) error terms. These predictions are at the dashboard-week level. We construct the marginal effects using the subsample of dashboard-weeks from 2018 after GDPR enforcement; we denote the set of post-GDPR weeks by W_{post}. We then compute the average marginal effect (AME) as follows:

$$AME = \frac{1}{N}\frac{1}{|W_{post}|}\sum_{i=1}^{N}\sum_{w \in W_{post}} \frac{y_{iw}^{treated} - y_{iw}^{untreated}}{y_{iw}^{untreated}} \tag{33}$$

where $|W_{post}|$ denotes the number of post-GDPR weeks in the sample and N is the number of dashboards in the sample.

D Testing the privacy frictions mechanism

Websites are reluctant to use obtrusive consent dialogs that add friction to the user's browsing experience. The Irish data regulator notes a "general resistance" among sites to introduce privacy frictions through the consent interfaces (Data Protection Commission, 2020b). For this reason, websites experiment with different consent interfaces to reduce privacy frictions and ensure a high consent rate (Long, 2020). Past research suggests that interruptions can hurt online usage (Lambrecht et al., 2011) and noted the long time required to read websites privacy policies (McDonald et al., 2009). Thus, the specific concern of privacy frictions is that they interrupt the user's browsing and deter users from continuing to browse the site—leading to a decrease in real outcomes. This motivates a simple empirical test: we expect the share of visits where the user *bounces* (browses a single page before leaving) will increase as privacy frictions increase.

We test for a privacy frictions mechanism by examining the effect of GDPR on bounce rates. Bounce rates are defined as the share of site visits with only a single pageview. More generally, bounce rates are a key diagnostic outcome in site analytics. Sites may wish to reduce bounce rates in order to increase ad and e-commerce revenue opportunities. On the other hand, high bounce rates can also indicate that the website effectively communicates information to the user: e.g., the temperature in Paris. In the pre-GDPR period, bounce rates average 42% across all dashboards and 37% in the e-commerce sample.

Bounce rates are useful because they indicate how dashboards seek consent. We expect that bouncing users are unlikely to provide explicit consent, because these users minimally interact with the site. If sites employ opt-out consent, then we expect bounce rates to rise somewhat if the privacy frictions mechanisms holds. If instead the site uses a strict opt-in model, then we expect to see a large drop in bounce rates because we expect few bouncing users will opt in to data recording. Comparing individual dashboard means across the pre- and post-GDPR periods suggests that the large majority of our sites see minimal changes to their bounce rates. In particular, only 3 sites exhibit patterns that are consistent with a strict, opt-in approach in that their bounce rates decrease more than 20 percentage points, but their recorded pageviews fall more than 50%. Thus, the vast majority of websites in our sample appear to employ an opt-out approach for consent, which is consistent with large surveys of website behaviors during that period (Sanchez-Rola et al., 2019; Utz et al., 2019; Johnson et al., 2022). As such, consistent with opt-out consent, we expect bounce rates to rise if privacy frictions are contributing substantially to our point estimates. We use bounce rates to test for the privacy frictions mechanism whereby GDPR consent dialogs dissuade users from further browsing.

We test for a change in bounce rates by reestimating our panel differences model (equation 3) with bounce rates as the dependent variable. Table 4 columns (1) and (2) present the model estimates: -0.275 percentage points (s.e. 0.304) for all sites and -0.354 percentage points (s.e. 0.591) for e-commerce sites. Thus, we find no statistically significant evidence that bounce rates change due to the GDPR and these null effect estimates are precisely estimated. This finding may allay website concerns about the privacy frictions mechanism after the GDPR, though more obtrusive consent dialogs that seek opt-in consent could create greater browsing frictions. We proceed under the assumption that the privacy frictions

Table 15: GDPR effect estimates: bounce rate

	(1)	(2)
Sample	All Dashboards	E-commerce
Dependent variable[†]	Bounce rate (percentage points)	
$\mathbb{1}\{2018\}\times\mathbb{1}\{\text{Post GDPR}\}$	-0.275	-0.354
	(0.304)	(0.591)
$\mathbb{1}\{2018\}$	0.311	0.111
	(0.304)	(0.539)
RSID + Week FE	Y	Y
R^2	0.861	0.790
N	69,344	22,436

Note: Standard errors clustered at the Dashboard + Week level. [†]Recorded outcomes.

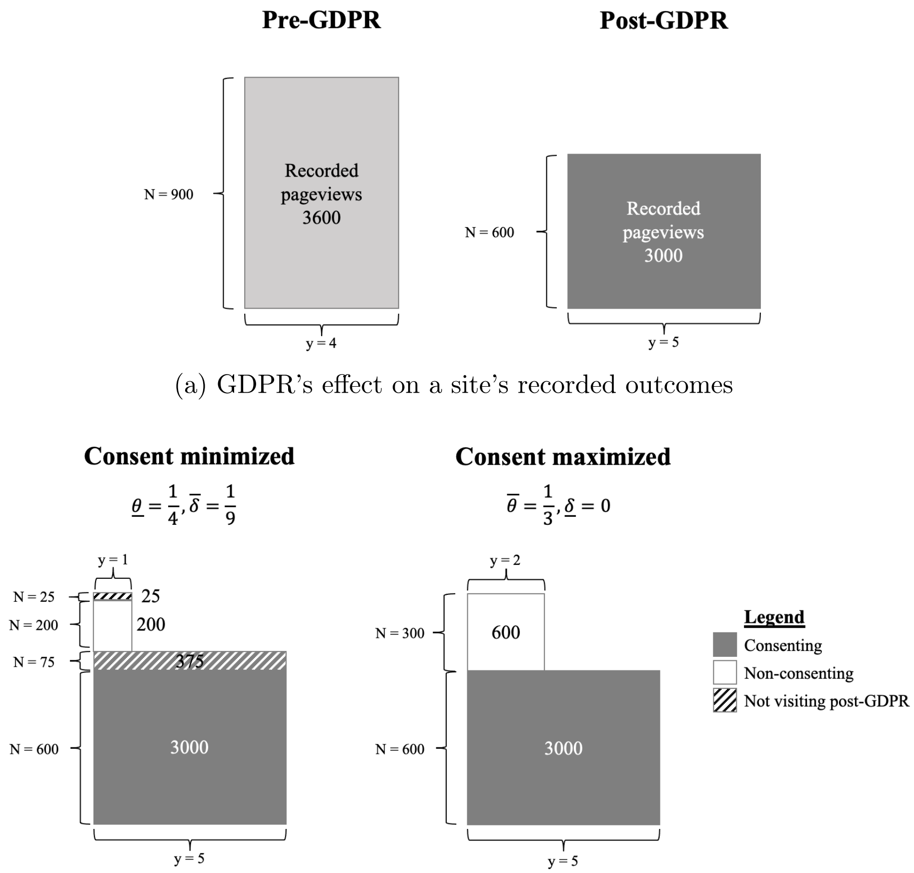
mechanism does not materially contribute to our estimated GDPR effect.

E A bounding example

Figure 9 illustrates our bounding approach from Section 5 using a numerical example. Figure 10a presents a simple example of a site that sees its recorded pageviews fall from 3,600 to 3,000 after the GDPR. The fall in recorded pageviews is due to changes in both recorded visits and recorded pageviews per visit. In this example, visits fall from 900 to 600, illustrated by the decrease in the height of the rectangle. As in our data, recorded pageviews per visit, rise post-GDPR from 4 to 5 pageviews. This change is reflected in the increase the rectangles width post-GDPR. Our approach decomposes visits into four types: 1) visits that would remain post-GDPR and provide consent; 2) visits that would remain, but refuse consent; 3) visits that leave post-GDPR, but would consent; and 4) visits that would leave post-GDPR and refuse consent. Fundamentally, we assume that we observe the sum of all four types pre-GDPR, and that the post-GDPR outcomes represent the first type of visits alone. As such, our bounding approach seeks to allocate the difference in recorded outcomes—here, 300 recorded visits and 600 recorded pageviews—to the remaining three user types.

Figure 10b illustrates our bounding approach. The right-hand side of Figure 10b shows the simplest case where the consent effect of the GDPR is maximized under the assumption that the GDPR has no real effect on visits. Consenting users are represented by the lower dark rectangle, while the white square represents non-consenting users. In this case, all the missing 300 visits and 600 pageviews are attributed to non-consenting visits (i.e., type two visits). This implies that non-consenting users are adversely selected, consuming only 2 pageviews per visit. The left-hand side of Figure 10b shows the more complicated case in which the consent effect is minimized and the real effect is maximized. Here, we use

Figure 9: Illustrative example of mechanism identification



(a) GDPR's effect on a site's recorded outcomes

(b) Corresponding bounds on GDPR's consent and real effects

Note: This figure presents a graphical intuition for our bounding analysis in Section 5.

the natural lower bound of 1 pageview per visit, which constrains the usage, conditional on visiting, of types two through four. Moreover, Assumption 4 holds that the missing visits, due to the GDPR's real harm, are missing at random with respect to consent. Given these assumptions, our bounds suggest that the pure consent effect represents 200 pageviews (type two visits). The real effect of the GDPR accounts for 400 pageviews: 375 pageviews from type three visits and 25 pageviews from type four visits.

F Bounds estimation

Section 5.1 derives tractable moment conditions, at the level of a dashboard j, which relate empirical moments to bounds on the underlying parameters of interest, $\Theta = \{\underline{\delta}, \bar{\delta}, \underline{\theta}, \bar{\theta}\}$. To begin to estimate these moments, we need empirical equivalents of the quantities $E\left[y_j^{obs}(Z_j)|V_j^{obs}(Z_j) = 1\right]$

and $E\left[V_j^{obs}=1|Z_j\right]$ with and without the GDPR, i.e., for $Z\in\{0,1\}$. Also, as we note in Section 5.1, we observe visits rather than the average visit probability $E\left[V_j^{obs}=1|Z_j\right]$. However, note that Corollaries 1 and 2 only require the ratio $\frac{E\left[V^{obs}|Z=1\right]}{E\left[V^{obs}|Z=0\right]}$ and equation (6) suggests that this equals the equivalent ratio for visits:

$$\frac{E\left[V_j^{obs}|Z_j=1\right]}{E\left[V_j^{obs}|Z=0\right]}=\frac{E\left[N_j^{obs}|Z_j=1\right]}{E\left[N_j^{obs}|Z_j=0\right]}$$

A final complication arises because we do not directly observe the usage and visits counterfactual outcomes; rather, we only observe either $Z=0$ or $Z=1$ at a point in time. In particular, simple averages before and after the GDPR are insufficient, as they may conflate seasonal variation with the GDPR's effect. To overcome this, we use the panel differences identification strategy introduced in Section 4. For each dashboard j and each outcome (y^{obs} and V^{obs}), we estimate a GDPR treatment effect β_j via the dashboard level analogue of equation (3). That is,

$$E\left[y_j^{obs}|V_j^{obs}=1,Z_j=0\right]=\frac{1}{T_1}\sum_{t\in T_1}y_{jt}^{2018}$$
$$E\left[y_j^{obs}|V_j^{obs}=1,Z_j=1\right]=\beta_j^y+\frac{1}{T_1}\sum_{t\in T_1}y_{jt}^{2018}$$
$$\frac{E\left[N_j^{obs}|Z_j=1\right]}{E\left[N_j^{obs}|Z_j=0\right]}=\beta_j^N$$

where T_1 denotes the total time observations from the pre-GDPR period of our data.

Then, we use the moment conditions from Corollaries 1 & 2 to estimate Θ. Specifically, we input the above quantities into equations (12-14) to define our estimation equations:

$$\begin{aligned}
g_{1j}\left(y_j^{2018},\beta_j,\Theta\right)=&\underline{\theta}\left(\beta_j^y+\frac{1}{T_1}\sum_{t\in T_1}y_{jt}^{2018}-\underline{y}\right)+\beta_j^y+\varepsilon_{1j}\\
g_{2j}\left(y_j^{2018},\beta_j,\Theta\right)=&1-\bar{\delta}-\frac{\beta_j^N}{(1-\underline{\theta})}+\varepsilon_{2j}\\
g_{3j}\left(y_j^{2018},\beta_j,\Theta\right)=&1-\overline{\theta}-\beta_j^N+\varepsilon_{3j}
\end{aligned}$$

with the moment condition that $E\left[g\left(y_j^{2018},\beta_j,\Theta\right)\right]=0$. Note that $\underline{\delta}=0$ is already determined by equation (15). This yields common Θ estimates across all dashboards, which we estimate using generalized method of moments (GMM). Note that our marketing effect lower bound estimation (Section 5.3) proceeds analogously using equations (16-18).

Conceptually, the ε_j terms relax the requirement that equations (12-14) hold exactly at the dashboard level. We do not directly observe the requisite counterfactual data and our dashboard-level estimates β_j are very imprecise. Instead, we are imposing that these relations hold on average across dashboards. Note that cases where the dashboard level estimates contradict our model—because estimate visits rise or usage falls post-GDPR—are exceptions and are captured by the ε_j terms. The advantage to this approach is that it allows for more heterogeneity in dashboard-level GDPR treatment effects. Another is that we do not impose cross-sectional restrictions on the distributions of dashboard size (i.e., number of visits N_j) and usage metrics (y_j).

G Proofs

Proposition 1

Part A

For part A), we start with the observed usage pre-GDPR $E\left[y^{obs}|V^{obs}=1, Z=0\right]$. Using the potential outcomes notation, we have

$$E\left[y^{obs}|V^{obs}=1, Z=0\right] = E\left[y^{obs}\left(0\right)|V^{obs}\left(0\right)=1, Z=0\right]$$

By the definitions of the recorded outcomes then Assumption 2, we have

$$\begin{aligned} E\left[y^{obs}\left(0\right)|V^{obs}\left(0\right)=1, Z=0\right] &= E\left[y\left(0\right)|V\left(0\right)=1, Z=0\right] \\ &= E\left[y|V\left(0\right)=1\right] \end{aligned}$$

Now, we apply the law of total expectation, so that we have

$$\begin{aligned} E\left[y|V\left(0\right)=1\right] = {} & E\left[y|V\left(0\right)=1, C\left(1\right)=0\right]\Pr\left[C\left(1\right)=0|V\left(0\right)=1\right] \\ & + E\left[y|V\left(0\right)=1, C\left(1\right)=1\right]\Pr\left[C\left(1\right)=1|V\left(0\right)=1\right] \end{aligned}$$

Using Assumption 3, we have

$$\begin{aligned} E\left[y|V\left(0\right)=1\right] = {} & E\left[y|V\left(0\right)=1, C\left(1\right)=0\right]\Pr\left[C\left(1\right)=0\right] \\ & + E\left[y|V\left(0\right)=1, C\left(1\right)=1\right]\Pr\left[C\left(1\right)=1\right] \end{aligned}$$

and by the definition of θ we have

$$\begin{aligned} E\left[y|V\left(0\right)=1\right] &= E\left[y|V\left(0\right)=1,C\left(1\right)=0\right]\cdot\theta \\ &\quad + E\left[y|V\left(0\right)=1,C\left(1\right)=1\right]\cdot\left(1-\theta\right) \end{aligned}$$

Now using Assumption 4, we have

$$\begin{aligned} E\left[y|V\left(0\right)=1\right] &= E\left[y|V\left(0\right)=1,C\left(1\right)=0\right]\cdot\theta \\ &\quad + E\left[y|V\left(0\right)=1,V\left(1\right)=1,C\left(1\right)=1\right]\cdot\left(1-\theta\right) \end{aligned}$$

We now show that $E\left[y|V\left(0\right)=1,V\left(1\right)=1,C\left(1\right)=1\right]$ is equivalent to observed usage post-GDPR. By Assumptions 1 then 2, we have

$$\begin{aligned} E\left[y|V\left(0\right)=1,V\left(1\right)=1,C\left(1\right)=1\right] &= E\left[y|V\left(1\right)=1,C\left(1\right)=1\right] \\ &= E\left[y|V\left(1\right)=1,C\left(1\right)=1,Z=1\right] \end{aligned}$$

by the definition of recorded outcomes, we have

$$E\left[y|V\left(1\right)=1,C\left(1\right)=1,Z=1\right]=E\left[y^{obs}|V^{obs}\left(1\right)=1,Z=1\right]$$

and simplifying to drop the potential outcomes we have

$$E\left[y^{obs}|V^{obs}\left(1\right)=1,Z=1\right]=E\left[y^{obs}|V^{obs}=1,Z=1\right]$$

Putting this all together and expressing in terms of observables, we have

$$\begin{aligned} E\left[y^{obs}|V^{obs}=1,Z=0\right] &= E\left[y\left(0\right)|V\left(0\right)=1,C\left(1\right)=0\right]\cdot\theta \\ &\quad + E\left[y^{obs}|V^{obs}=1,Z=1\right]\cdot\left(1-\theta\right) \end{aligned}$$

solving for θ obtains the equation in Proposition 1A).

Part B

We start by examining the change in observed visits due to the GDPR:
$E\left[V^{obs}=1|Z=1\right]-E\left[V^{obs}=1|Z=0\right]=E\left[V\left(1\right)\cdot C\left(1\right)=1|Z=1\right]-E\left[V\left(0\right)=1|Z=0\right]$

Where the equality follow from the definition of V^{obs}. Then, given Assumption 3 we have:

$$E\left[V^{obs}=1|Z=1\right]-E\left[V^{obs}=1|Z=0\right]=E\left[V\left(1\right)=1\right]E\left[C\left(1\right)=1\right]-E\left[V\left(0\right)=1\right]$$

Substituting in equations (10) and (9) on the right hand side yields:

$$E\left[V^{obs}=1|Z=1\right]-E\left[V^{obs}=1|Z=0\right]=E\left[V\left(0\right)=1\right]\cdot\left(\left(1-\delta\right)\left(1-\theta\right)-1\right)$$

Finally, we substitute $E\left[V\left(0\right)=1\right]=E\left[V^{obs}=1|Z=0\right]$ and simplify to get:

$$\delta=1-\frac{E\left[V^{obs}=1|Z=1\right]}{E\left[V^{obs}=1|Z=0\right]\left(1-\theta\right)}$$

Proposition 2

Part A

Part A follows from the proof of Corollary 2.

Part B Under Assumption 5, the consent rate estimated in Proposition 2 Part A serves as a lower bound for the Email and Display attribution channels. Then, Part B follows by applying Corollary 1.